

Tytuł: **Wykorzystanie metod sztucznej inteligencji w badaniach nad umysłem**

Autor: Adam Chuderski / [adam.chuderski@emapa.pl](mailto:adam.chuderski@emapa.pl)

Recenzent: prof. dr hab. Jan B. Gajda (UŁ), dr Marek Miszczyński (UŁ)

Źródło: <http://www.kognitywistyka.net> / [mjkasperski@kognitywistyka.net](mailto:mjkasperski@kognitywistyka.net)

Data publikacji: 06 III 2002

## 1. Krótka historia naukowego badania umysłu

Pierwsze rozważania o funkcjonowaniu obiektu, który obecnie zwiemy umysłem, tak jak i w ogóle rozważania o naturze wszechświata, poczęto czynić w starożytnej Grecji. Tam, bowiem narodziła się filozofia – pierwsze systematyczne i ogólne dociekania człowieka o sobie samym i otaczającym go świecie. Za pomocą filozoficznej spekulacji dwaj najwybitniejsi filozofowie greccy: Platon w dialogu *Fajdros* i Arystoteles w *Traktacie o duszy* rozważali te same problemy dotyczące procesów poznawczych człowieka, które do dzisiaj próbuje rozwiązać nauka. Podobnie jak dwa i pół tysiąca lat temu Grecy, tak i my teraz próbujemy dowiedzieć się jaką wiedzą człowiek dysponuje od urodzenia, jak pozyskuje doświadczenie, jak postrzega świat, jaka jest wzajemna relacja między intelektem (umysłem), wolą i emocjami. Do połowy XIX w. wiedza o psychice człowieka obejmowała niewiele więcej poza tym, co odkryli starożytni filozofowie.

Wraz z początkiem XX w. psychologia, oddzielając się od filozofii, ukonstytuowała się jako samodzielna nauka. Powstały różne szkoły psychologiczne próbujące naukowo odpowiedzieć na pytanie, jak funkcjonuje umysł człowieka. Początkowo stosowano metody badawcze oparte na introspekcji – wglądzie człowieka we własne stany umysłowe. Metodzie tej brakowało jednak niezbędnego w nauce obiektywizmu. Obserwowane zjawisko było bowiem subiektywnym przeżyciem badającego i badanego w jednej osobie.

Radykalne rozwiązanie pojawiło się w 1913 r., wraz z publikacją artykułu Jamesa Watsona *Psychologia, jak ją widzi behawiorysta* (Watson 1913/1990). Artykuł ten zapoczątkował nową szkołę psychologiczną zwaną behawioryzmem, wprowadzającą do psychologii rygorystyczną metodę badawczą. Wg Watsona stany umysłu człowieka nie mają wpływu na jego zachowanie. Nie potrzebujemy więc pojęcia umysłu w psychologii. Wystarczy badać bodźce działające na człowieka i reakcje, jakie powodują. Powtarzające się występowanie określonego bodźca w środowisku powoduje ustalenie się jednej, odpowiadającej mu reakcji człowieka. Powstanie nawyk, czyli uporządkowana para <bodziec, reakcja>. Struktura wszystkich takich par określa osobowość człowieka, a znajomość tej struktury pozwala przewidzieć jego zachowanie. Następcy Watsona, m.in. Skinner (Skinner 1971) próbowali wyjaśnić nawet tak skomplikowane zachowania jak użycie języka. Proponowali też budowę idealnego społeczeństwa poprzez wykształcenie w jego członkach odpowiednik nawyków. Zakładali bowiem, że osobowość człowieka daje się zmienić w sposób doskonały, jeśli tylko

użyje się odpowiednich sekwencji bodźców i kształtujących odpowiednie reakcje kar i nagród.

Dominujący w latach 30-tych i 40-tych behawioryzm, mimo iż stanowił spójną teorię oraz wykorzystywał ściśle i obiektywne metody badawcze, poddany został poważnej krytyce. Stracił aktualność, zastąpiony przez dwie stworzone w latach 50-tych teorie, uznające umysł, jego stany i procesy za podstawowy klucz do wyjaśnienia zagadki zachowania się człowieka. Owe teorie to lingwistyka Chomsky'ego oraz nowy nurt w psychologii, zwany psychologią poznawczą.

Psychologia poznawcza<sup>1</sup> wyjaśnia zachowanie człowieka nie tylko poprzez działające nań bodźce, ale także poprzez stan jego umysłu. Traktuje aktywność umysłową człowieka i wynikające z niej zachowanie w kategoriach przetwarzania informacji. Człowiek otrzymuje i sam odbiera informację z zewnątrz, interpretuje ją w zależności od informacji już posiadanej, przetwarza i generuje nową informację, dzięki której podejmuje działanie. Proces ten składa się z etapów, różniących się przebiegiem procesów przetwarzania danych: w percepcji odbiera i wstępnie przetwarza informację, podczas myślenia przekształca ją w celu rozwiązania napotkanego problemu, następnie podejmuje decyzję skutkującą podjęciem określonego zachowania (skutek behawioralny) lub powstaniem nowych struktur w umyśle (skutek poznawczy). Zdolność do generowania nowych struktur informacyjnych jest ważną cechą poznawczej koncepcji człowieka. Zarówno otrzymywane, jak i generowane informacje mogą stać się stałymi składnikami wiedzy, jaką człowiek posiada. Te struktury to. m.in. pojęcia, idee, wyobrażenia, cele, pragnienia itp. Wrodzone zdolności przetwarzania informacji oraz uzyskane w trakcie życia osobniczego struktury informacyjne (reprezentacje świata oraz obraz siebie) decydują, wg psychologów poznawczych, o osobowości człowieka. Już w początkach rozwoju psychologii poznawczej proces przetwarzania informacji zaczęto ujmować ilościowo. Jedną z najśłynniejszych prac określających ilość informacji przetwarzanej przez konkretny składnik umysłu – pamięć krótkotrwałą – była praca G. Millera *Magiczna liczba siedem plus minusz dwa...* (Miller 1956). W pracy tej postulowano, przyjmowaną z niewielkimi modyfikacjami i dzisiaj, tezę o pojemności pamięci krótkotrwałej człowieka wynoszącej  $7 \pm 2$  elementy. Ponieważ opierano się na aparacie teorii informacji, początkowo nie uwzględniano treści, a tylko ilość informacji. Kilkanaście lat później rozpoczęto też jakościową analizę przetwarzania informacji w umyśle, wyodrębniono pewne struktury, badano ich zawartość i operacje, którym ta zawartość podlega.

Drugą z teorii, która postulowała istotny wpływ istniejących w umyśle struktur na zachowanie, była lingwistyka Noama Chomsky'ego. Analizowana w niej użycie i recepcję mowy oraz, wtórnego wobec mowy, pisma. Chomsky zauważył, że człowiek jest zdolny do twórczego generowania zdań, jakich nigdy przedtem nie słyszał. Wyjaśnienie behawioralne, mówiące iż generowanie zdań odbywa się na zasadzie nawyku wiążącego pasujące do bodźca przyswojone wcześniej zdanie, Chomsky uważał za nieadekwatne. Przedstawił pogląd, że w umyśle muszą istnieć wrodzone reguły umożliwiające produkcję nieskończonej liczby zdań na podstawie skończonego zasobu przyswojonych przez człowieka słów. Wyodrębnienie owych reguł pozwoliłoby na sformalizowanie procesów językowych. Lingwistyka była więc kolejnym krokiem w kierunku 'informatyzacji' procesów poznawczych człowieka (Lyons 1998).

<sup>1</sup> Termin ten występuje również jako 'psychologia kognitywna'. Przep. M. Kasperski.

Przenieśmy się teraz w zupełnie inne miejsce w hierarchii nauk, gdzie rozwinęła się w połowie XX w. technologia budowy komputerów i teoretyczna wiedza o ich funkcjonowaniu. Człowiek od kilkuset lat próbował stworzyć maszyny realizujące pewne funkcje umysłowe człowieka. Najpierw próbowano zautomatyzować liczenie, jako że matematyka doskonale wyjaśniała jakich operacji należy użyć, aby odpowiednio przekształcając wprowadzone do maszyny liczby otrzymać prawidłowy wynik. Około roku 1650 francuski myśliciel Blaise Pascal zaprojektował i wykonał mechaniczny kalkulator zdolny do dodawania i odejmowania liczb. Maszyna składała się z systemu kół zębatych, na których grawerowano cyfry. Ustawiając część kół w położeniu odpowiadającym liczbom na wejściu i uruchamiając korba mechanizm kalkulatora użytkownik uzyskiwał żądany wynik dodawania lub odejmowania. Mnożenie można było wykonywać jedynie poprzez powtarzanie operacji dodawania (Pratt 1987, s. 51).

Zdolność mnożenia posiadały już mechaniczne kalkulatory wykonane według projektu niemieckiego filozofa Gottfrieda Leibniza (1685 r.). Przeszedł on jednak do historii informatyki nie jako projektant kalkulatorów, ale jako pomysłodawca maszyny automatyzującej wnioskowanie – protoplasty nowoczesnych systemów sztucznej inteligencji. Podstawą projektu tej maszyny miał być uniwersalny język logiczny zwany *characteristica universalis*<sup>2</sup>. W tym języku dałoby się, wg Leibniza, przeprowadzać mechanicznie metodę rachunkową zwaną *calculus universalis*, pozwalającą rozwiązywać wszelkie problemy naukowe i filozoficzne (Murawski 1995, ss. 48-49). Oznaczając np. pojęcie ‘człowiek’ symbolem *AB*, a ‘człowiek rasy czarnej’ – *ABC* wystarczyło tylko sprawdzić, czy pojęcie *ABC* daje się rozłożyć tak, aby uzyskać symbol *AB*. Interpretując *AB* jako znak pojęcia ‘człowiek’ można już orzec, że pojęcie ‘człowiek rasy czarnej’ zawiera się w pojęciu ‘człowiek’. Maszyna realizująca *calculus universalis* nie została przez Leibniza stworzona, lecz doniosłość jego idei jest historyczna. Po raz pierwszy stworzono projekt zapisu myśli za pomocą systemu symboli, które swoje znaczenie zyskują dzięki interpretacji oraz wskazano na konieczność istnienia reguł, umożliwiających przekształcanie tych symboli<sup>3</sup>. System symboli miał być ogólny (pozwalając na zapis całej wiedzy) i ścisły. Dzięki maszynie Leibniza dwóch myślicieli zamiast spierać się w nieprecyzyjnym języku naturalnym miałyby zawołać „*Calculemus!*” (*Porachujmy!*), zakodować problem w maszynie i uzyskać rozwiązanie. Idea ta stanowi do dziś fundament teorii i praktyki sztucznej Inteligencji.

W XVIII i XIX w. następował szybki rozwój mechanicznych maszyn liczących, na które istniało olbrzymie zapotrzebowanie ze strony rosnącego w siłę przemysłu, handlu i administracji. Rozwój ten znalazł ukoronowanie w pracy angielskiego matematyka i inżyniera Charlesa Babbage’a. Zaprojektował on i nadzorował wykonanie w 1832 r. maszyny zwanej *Difference Engine*, która umożliwiała obliczanie wartości pochodnych wielomianów do trzeciego stopnia. Maszyna drukowała też wyniki swojej pracy. Babbage zaprojektował także potężniejszą wersję *Difference Engine*, której jednak z powodów finansowych i organizacyjnych nie dokończył. Maszynę tę, potrafiącą obliczać wartości pochodnych 6-tego stopnia, wykonał w 1853 r. szwedzki konstruktor George Scheutz (Pratt 1987, ss. 102-110).

<sup>2</sup> Tutaj Leibniz częściowo opierał się na dokonaniach konstruktorów języków uniwersalnych, zwłaszcza Rajmunda Lullusa (1234-1315) i jego dziele *Ars magna generalis et ultima* (tj. *Sztuka wielka ogólna i najwyższa*). Warto podkreślić, że Lullus, tak samo jak Leibniz, był konstruktorem maszyny matematyczno-logicznej. Por. w tej sprawie: M. Jurkowski, *Od wieży Babel do języka kosmitów. O językach sztucznych, uniwersalnych i międzynarodowych*, KAW, Białystok 1986, ss. 17, 30-35. Przyp. M. Kasperski.

<sup>3</sup> Zdaje się, że nie do końca tak to było! Pierwszy taki system stworzył już wspomniany przeze mnie R. Lullus. Jednakże w większości literatury nawet nie wspomina się o jego istnieniu stąd powołanie się na pierwszeństwo Leibniza. Przyp. M. Kasperski.

Choć maszyna Babbage'a dawała nowe możliwości obliczeniowe, prawdziwie rewolucyjne rozwiązania konstruktor zawarł w nie zrealizowanym nigdy do końca projekcie maszyny pozwalającej wykonywać wszystkie znane działania algebraiczne, zwanej *Analytical Engine*. Rewolucyjność pomysłu polegała na tym, że maszyna ta miała być programowalna. Za pomocą kart dziurkowanych<sup>4</sup> możliwe byłoby wprowadzenie do maszyny sekwencji operacji algebraicznych, które miałyby następnie wykonać. Aby to umożliwić, Babbage zaprojektował drugie nowe rozwiązanie – pamięć, w której maszyna przechowywałaby pośrednie wyniki działań. Teoretycznie maszyna ta pracowałaby analogicznie do dzisiejszych komputerów, lecz tworzywo użyte do jej realizacji – metalowe tryby, koła i pręty – stanowiło nieprzekraczalne ograniczenie w jej rozwoju (Pratt 1987, ss. 114-127).

Ograniczenie to pokonano dzięki rozwojowi elektroniki w czasie II wojny światowej. W 1942 r. zbudowano ENIAC-a (*Electronical Numerical Integrator and Computer*), maszyną opartą na lampach elektronowych i składającą się z zestawu jednostek, z których każda realizowała specyficzne jej działania. Maszyna ta przechowywała wyniki obliczeń w postaci elektronicznej, nie dawała jednak pełnej możliwości programowania. Aby zmienić realizowaną przez ENIAC-a funkcję trzeba było ingerencji w jego fizyczną strukturę, należało zmienić układ kabli i poprzestawiać odpowiednie przełączniki. Szybko jednak zaczęły powstawać maszyny umożliwiające doskonalszą zmianę programu. Dzięki wysiłkom matematyka Johna von Neumanna powstała IAS Machine, prototyp dzisiejszych komputerów. Maszyna przechowywała w tej samej pamięci zarówno dane jak i program, opierała się na kodzie binarnym i dysponowała pamięcią zewnętrzną. W końcu lat 50-tych lampy zastąpiono tranzystorami i rozpoczął się wyścig producentów komputerów w uzyskiwaniu coraz większej szybkości obliczeniowej, pojemności pamięci i mniejszych rozmiarów sprzętu (miniaturyzacji) (Pratt 1987, ss. 162-171).

Wykorzystując osiągnięcia rozwijającej się burzliwie w I poł. XX w. logiki formalnej stworzono także teoretyczne podstawy programowania komputerów. Szczególny wkład, opisany w rozdziale czwartym niniejszej pracy, wnieśli wspomniany John von Neumann oraz angielski logik i matematyk Alan Turing. Drugi z wymienionych naukowców znany jest, oprócz swojej teorii algorytmów, także z rozważań, czy maszyna wyposażona w odpowiedni program może myśleć. W swojej historycznej pracy *Can a Machine Think* (Turing 1950) autor zauważa, że ze względu na brak jednoznacznej definicji słowa 'myśleć', odpowiedź na pytanie: czy maszyna może myśleć? zależy od tego, co odpowiadający pod pojęciem 'myślenie' rozumie. Aby uniknąć wieloznaczności proponuje on wprowadzenie prostej procedury rozstrzygającej. W procedurze tej, zwanej testem Turinga, uczestniczą trzy obiekty: człowiek-sędzia, maszyna i drugi człowiek. Człowiek-sędzia, pozbawiony kontaktu wzrokowego z maszyną i człowiekiem [drugim] zadaje pytania, posługując się terminalem. W zależności od odpowiedzi orzeka, który z dwóch badanych obiektów jest maszyną. Człowiek stara się pomóc sędziemu w poprawnym sądzie, maszyna może wprowadzać go w błąd. Jeżeli odpowiednio duża liczba ludzi uczestniczących w teście jako sędzia nie rozpozna odpowiedzi maszyny, mamy prawo – wg Turinga – uznać, że maszyna ta myśli. W swojej pracy Turing wyraża wiarę, że maszyna będzie w stanie rozwiązać problemy właściwe człowiekowi, choć przedstawia i dyskutuje wiele zarzutów, jakie taki pogląd może napotkać.

Poparta teoretycznymi rozważaniami Turinga zaczyna rozwijać się teoria i praktyka sztucznej inteligencji. Dysponując dosyć już rozwiniętymi komputerami pierwsi badacze sztucznej inteligencji: John McCarthy i Marvin Minsky na MIT, Herbert Simon i Alan Newell na

---

<sup>4</sup> Kart perforowanych. Przep. M. Kasperski.

Canegie Institute of Technology (obecnie Carnegie-Mellon University), oraz inni badacze skupieni głównie wokół tych dwóch ośrodków, rozpoczęli prace nad programami, które wykonywały zadania, zwykle wymagające inteligencji człowieka. Ukonstytuowała się nazwa dla klasy tych programów – ‘*Artificial Intelligence*’ (AI), zaproponowana po raz pierwszy przez McCarthy'ego jako nazwa konferencji odbywającej się w 1956 r. w Dartmouth College, New Hampshire. Odnotowano pierwsze sukcesy: program do gry w warcaby stworzony przez Samuela, program do rozpoznawania obrazów Selfridge'a oraz dwa systemy autorstwa Newella i Simona: program do automatycznego dowodzenia twierdzeń logicznych nazwany Teoretykiem Logiki oraz najważniejsze osiągnięcie tego okresu – *General Problem Solver* (GPS) (Pratt 1987, rozdz. 14 i 15).

Program ten miał za zadanie za pomocą odpowiedniego wnioskowania rozwiązywać wszystkie dające się odpowiednio zakodować problemy. Niósł on kilka rewolucyjnych rozwiązań, które wpływają do dzisiaj na rozwój nauki o sztucznej inteligencji. Po pierwsze, autorzy chcieli uzyskać duży stopień podobieństwa działania programu do procesów podejmowania decyzji przez człowieka. Struktura GPS uwarunkowana była więc nie tylko pomysłami programistów, ale także badaniami psychologów. Po drugie, miał to być program ogólny, przeznaczony do rozwiązywania szerokiej klasy problemów. Po trzecie, GPS wykorzystywał w swoim działaniu cele (które należy osiągnąć), symbole (zwane obiektami, które można przekształcać) i operatory (służące przekształcaniu symboli) (Newell i Simon 1961). Stanowi to do dziś wzorcową strukturę programów modelujących/realizujących sztuczną inteligencję. Nazwa programu była zbyt szumna, zważywszy na jego skromne możliwości, ale dobrze oddawała entuzjazm pierwszych lat *Artificial Intelligence*.

W latach 60-tych budowano wiele systemów modelujących procesy umysłowe człowieka, a pojęcia oraz metody nauki o komputerach i sztucznej inteligencji przenikały do psychologii poznawczej. Zaczęto zauważać odpowiadające sobie pojęcia w obu naukach: myśl – program, struktury umysłowe – struktury danych, procesy psychiczne – algorytmy. Zaczęto postulować konieczność stworzenia teorii, która opisywałaby zarówno funkcjonowanie umysłu człowieka, jak i systemów sztucznej inteligencji (Thagard 1987, s. 5). Podstawę filozoficzną takiej teorii stanowił nowy pogląd na problem wzajemnego stosunku umysłu do ciała, zwany problemem psychofizycznym<sup>5</sup>. Pogląd ten, zwany funkcjonalizmem, został stworzony przez amerykańskiego filozofa z uniwersytetu Harvarda – Hilarego Putnama<sup>6</sup>.

Przed Putnamem poglądy filozofów na to, czym jest umysł, odnosiły się do jego substancji. Upraszczając znacznie problem, poglądy te zawierały się pomiędzy dwoma skrajnościami:

1. Umysł jest jedyną istniejącą substancją, a to, co uważamy za materialne obiekty, to wyłącznie wrażenia tego umysłu (skrajny idealizm).
2. Istnieją jedynie obiekty materialne, a to, co uważamy za umysł można utożsamić z działaniem najbardziej skomplikowanego z nich – mózgu (skrajny materializm).

Z poglądów pośrednich warto wymienić teorie dualistyczne, których autorzy głosili, że mózg i umysł to dwie różne substancje, bądź wchodzące w interakcję ze sobą, bądź sterowane przez odrębną siłę. Putnam, zainspirowany komputerami, zaproponował zupełnie nowe ujęcie: umysł ma się do mózgu tak, jak *software* do *hardware* komputera. Umysł jest więc

<sup>5</sup> Z ang. *mind-body problem*. Przep. M. Kasperski.

<sup>6</sup> Jednakże, pod nieodzownym wpływem tekstu Turinga! To przecie Turing właśnie, proponując swój test [*imitation game*] na to czy maszyna myśli, proponuje rozważanie problemu umysłu i myślenia za pomocą badania jego funkcji. Przep. M. Kasperski.

programem mózgu. A zatem, podobnie jak nie ma sensu analiza programu poprzez badanie fizycznej budowy komputera, tak i umysł należy badać w oderwaniu od mózgu, jako obiekt abstrakcyjny. Istotna jest funkcja umysłu, należy więc rozpatrywać wyłącznie sposób jego działania. Od razu zauważono wniosek wynikający z przyjęcia funkcjonalizmu: jeśli programy nie są związane z jednym typem komputera, to i umysł nie musi wymagać do swego funkcjonowania mózgu. Jeśli stworzy się maszynę, do której udałoby się wprowadzić program działający tak, jak program mózgu, nie będzie różnicy między tymi dwoma programami. Program działający w maszynie również moglibyśmy nazwać umysłem. Wraz z powstaniem funkcjonalizmu naukowcy uzyskali uzasadnienie filozoficzne do konstrukcji dziedziny naukowej integrującej psychologię poznawczą (rozważającą umysł jako system przetwarzający informację) oraz naukę o sztucznej inteligencji (nadającą systemom przetwarzającym informację cechy umysłu).

Dyscyplina ta, stawiająca sobie za zadanie wyjaśnienie przebiegu procesów poznawczych korzystając z metod symulacji komputerowej, oficjalnie ukonstytuowała się w Stanach Zjednoczonych w 1975 r. Uzyskała nazwę – *cognitive science*, rami organizacyjne oraz pierwsze pieniądze na finansowanie badań – początkowo 15, a ostatecznie 20 milionów dolarów przeznaczonych przez Fundację im. Alfreda Sloana na kilkuletni program badawczy (Domańska 1991). W 1976 roku zaczęto wydawać kwartalnik pod tytułem "Cognitive Science" przedstawiający wyniki badań w tej dziedzinie. W 1979 roku powstało towarzystwo naukowe *Cognitive Science Society*, z siedzibą na uniwersytecie Michigan. Obecnie liczy ono ponad 1000 członków zwyczajnych oraz wielu członków afiliowanych i studentów. Od 1979 roku odbywają się także coroczne konferencje naukowe, na które zjeżdżają się naukowcy z całego świata. Na wielu uczelniach kształcą się studenci na kierunku *cognitive science*. W Polsce studia i seminaria w zakresie *cognitive science* prowadzone są m.in. na Uniwersytecie Mikołaja Kopernika i na Uniwersytecie Jagiellońskim.

## 2. Nauka poznawcza

W niniejszym rozdziale podjęta zostanie próba przedstawienia, czym zajmuje się *cognitive science*, poprzez przytoczenie kilku definicji tej dyscypliny, omówienie przedmiotu jej badań oraz relacji do innych dziedzin nauki. Najpierw jednak przedstawione będą funkcjonujące w polskiej literaturze tłumaczenia tej angielskiej nazwy.

W polskiej literaturze naukowej przyjęło się stosować kilka nazw dla denotacji *cognitive science*. Najpopularniejsze to: 'nauka poznawcza' (Maruszewicz 1996; Domańska 1991), 'nauka o poznawaniu' (Kurcz 1987), 'nauka o poznaniu' (Koziński 1996) i 'kognitywistyka' (Chlewiński 1999). Używa się także terminu nieprzetłumaczonego (Bobryk 1987; Bobryk 1994; Piłat 1994). W niniejszej pracy używana będzie nazwa 'nauka poznawcza'.

### 2.1. Czym zajmuje się nauka poznawcza?

W literaturze przedmiotu funkcjonuje wiele definicji nauki poznawczej. Pozwalają one zwięźle określić, czym zajmuje się ta dyscyplina, jednak przyjmują różne punkty odniesienia; metody, przedmiot badania czy też stosunek do innych nauk.

Poniższa definicja (Eysenck 1990) kładzie nacisk na wpływ innych nauk na badania prowadzone w obrębie nauki poznawczej:

Termin 'nauka poznawcza' odnosi się do interdyscyplinarnego studium dotyczącego nabywania i użycia wiedzy. W studium to wkład wnoszą: nauka o sztucznej inteligencji, psychologia, lingwistyka, filozofia, antropologia, neurofizjologia [*neuroscience*] i nauki o wychowaniu. Ruch ten odznacza się szerokim zasięgiem i zróżnicowaniem, zawierając w sobie kilka punktów widzenia. Nauka poznawcza rozwinęła się dzięki trzem osiągnięciom: wynalazkowi komputerów i próbach stworzenia programów wykonujących zadania, które właściwe są ludziom; rozwojowi psychologii poznawczej (...); oraz rozwinięciu się w dziedzinie lingwistyki teorii gramatyki generatywnej i teorii jej pochodnych. Nauka poznawcza jest syntezą zajmującą się istotą wiedzy, na jakiej opiera się ludzkie poznanie, procesami przetwarzania informacji przez człowieka i komputerowym modelowaniem tychże procesów. Istnieje pięć głównych pól badawczych w nauce poznawczej: reprezentacja wiedzy, język, uczenie się, myślenie i percepcja.

Podobna jest następująca definicja (Salo 1997):

Z historycznego punktu widzenia nauka poznawcza była całkiem nową propozycją w naukowym badaniu umysłu: był to pogląd, że ludzki umysł jest systemem manipulującym symbolami, pewnego rodzaju komputerem. Tak więc na początku nauka poznawcza była fuzją nauki o sztucznej inteligencji i psychologii poznawczej. (...) Obecnie, także lingwistyka i neurofizjologia [*neural sciences*] dołączają do owej fuzji, łącząc się coraz bardziej. Dołączają także inne działy psychologii. Filozoficzna debata na tych zagadnieniach jest gorąca, zróżnicowana i szersza niż kiedykolwiek, nauka o komputerach [*computer science*] dostarcza podstawy do weryfikacji teorii, nawet fizyka zaczyna być wykorzystywana!

Inna z definicji (Simon i Kaplan 1990) wskazuje, że to przedmiot badania, czyli systemy inteligentne, jest kluczem dla określenia czym jest nauka poznawcza:

[nauka poznawcza] ... jest to studium inteligencji i systemów inteligentnych [*intelligent systems*], ze szczególnym odniesieniem się do zachowania inteligentnego jako procesu obliczeniowego [*computation*].

Podobna definicja przedstawia się następująco (Bobryk 1994):

CS [czyli *cognitive science*] jest nauką o systemach (jednostkach, mechanizmach, urządzeniach), którym można przypisać zdolność myślenia i poznawania, czyli cechę inteligencji. Jak do tej pory CS zna dwie podstawowe klasy systemów inteligentnych: jednostki ludzkie (...) i odpowiednio zaprogramowane (przez co obdarzone inteligencją) komputery.

Poniższa definicja uwypukla fakt zdolności do przetwarzania informacji jako kluczowej cechy przedmiotu badań nauki poznawczej (Salo 1997):

Badacze w obrębie nauki poznawczej [*cognitive scientists*] postrzegają umysł ludzki jako system otrzymujący, przechowujący, odtwarzający, przetwarzający i przekazujący informację.

Kolejna definicja oprócz przedmiotu badań nauki poznawczej przedstawia także jej metodę badawczą (Domańska 1991):

Dyscyplina, która stawia sobie za zadanie wyjaśnienie przebiegu procesów poznawczych i czyni to, korzystając z metod symulacji komputerowej (...) Nauka poznawcza zajmuje się badaniem systemów poznawczych „w ogóle”, niezależnie od tego, czy występują u człowieka, czy w komputerze.

I wreszcie ostatnia definicja (Green i in. 1996):

Zdefiniujmy zakres nauki poznawczej jako interdyscyplinarne, naukowe studium umysłu. Praktyka i wiedza tej nauki są pochodną osiągnięć dyscyplin, które przyczyniły się do jej powstania: nauki o komputerach, lingwistyki, neurofizjologii [*neuroscience*], psychologii, neuropsychologii poznawczej i filozofii. Celem jej jest zrozumienie działania umysłu w kategoriach procesów manipulacji na reprezentacjach. Umysł, a więc i podstawa inteligentnego zachowania się w świecie, jest widziany w kategoriach obliczeń [*computations*] albo przetwarzania informacji [*information-processes*].

Lub krócej w tej samej pracy:

CS to interdyscyplinarna dyscyplina naukowa badająca umysł jako system przetwarzający informację.

Podsumowując powyższe definicje nauki poznawczej można stwierdzić, że jest ona interdyscyplinarną nauką zajmującą się badaniem systemów przetwarzających informację w sposób tak wysoce zorganizowany, że ich działanie może być nazwane inteligentnym. Systemy te określa się jako systemy poznawcze [*cognitive systems*]. Przetwarzanie informacji przez systemy poznawcze polega na manipulowaniu reprezentacjami, które odnoszą się do świata zewnętrznego. Nauka poznawcza zakłada, że proces ten daje się opisać w sposób formalny. Ma on charakter obliczeniowy [*computable*]. Dzięki temu, uzasadnione jest wykorzystywanie metod symulacji komputerowej do badania umysłu człowieka. Wykorzystaniu owych metod poświęcona jest większa część niniejszej pracy.

## 2.2. System poznawczy

Inteligentne zachowanie się systemu poznawczego jest trudne do zdefiniowania, ponieważ nie istnieje zadowalająca definicja inteligencji. Biorąc pod uwagę to, co wspólne istniejącym definicjom inteligencji, chodzi tu o cechę, która pozwala systemowi przystosować się do zmieniających się warunków i wykonywać nowe zadania; cechę, która implikuje zdolność uczenia się i efektywnego rozwiązywania problemów poprzez wnioskowanie, rozumienie, porównywanie itp. Powyższe cechy pociągają za sobą konieczność efektywnego przetwarzania informacji (Tomaszewski 1995, ss. 84-85).

System poznawczy to nazwa klasy systemów, do których należy zarówno umysł człowieka jak i odpowiednio zaprogramowane komputery. Klasa ta obejmie być może w przyszłości także ssaki naczelne, które wykazują pewne cechy zachowania inteligentnego, jak np. rozwiązywanie prostych problemów, użycie pewnych fragmentów języka migowego czy porównywanie liczebności zbiorów zawierających po kilka elementów.

Nauka poznawcza podczas badania funkcjonowania systemu poznawczego całkowicie abstrahuje od jego materii, substratu w jakim przebiegają procesy poznawcze. Ważna jest dla niej wyłącznie struktura i przebieg procesów poznawczych, które – wg zwolenników funkcjonalizmu – można badać w oderwaniu od ich fizycznego nośnika.

## 2.3 Nauka poznawcza a inne dyscypliny nauk

Spróbuję teraz porównać naukę poznawczą z innymi dziedzinami wiedzy o umyśle i mózgu, aby wykazać, że nauka poznawcza posiada własną metodę badawczą, różną od metod tychże dziedzin. Choć w ogromnym stopniu korzysta z ich dorobku, jej osiągnięcia są czymś więcej



niż tylko sumą ich wyników. Nauka poznawcza nie korzysta zarazem ze wszystkich danych nagromadzonych przez tradycyjne nauki (i filozofię umysłu), a tylko z tych, które mieszczą się w jej wizji umysłu jako systemu formalnego manipulującego reprezentacjami. Jak pisze Bobryk (Bobryk 1994) selekcja wyprzedza integrację i kontynuację, mając wyraźnie ukierunkowany charakter eliminuje wstępnie zarówno pewne dane zgromadzone przez wyjściowe dyscypliny, jak i pewne ich interpretacje.

Od nauki o sztucznej inteligencji naukę poznawczą odróżnia nacisk na eksperymenty i teorie, służące wyjaśnieniu mechanizmu funkcjonowania ludzkiego umysłu. Osiągnięcia nauki o sztucznej inteligencji wykorzystywane są jako aparat pojęciowy i metody symulacji pracy ludzkiego umysłu. Jednak podczas konstrukcji modeli umysłu naukowcy poznawczy nie wykorzystują bezpośrednio rozwiązań opracowanych przez badaczy AI, lecz starają się możliwie ściśle uzgodnić procesy realizowane przez komputer z tymi, które przebiegają w umyśle. Nauka o sztucznej inteligencji ma swobodę w znajdowaniu rozwiązań dla problemów, kładzie się tu nacisk na efektywność rozwiązania. W nauce poznawczej modele inteligencji mają działać tak, jak umysł ludzki, uwarunkowane są więc jego ograniczeniami. Znanych jest kilka skomplikowanych, lecz efektywnych metod automatycznego dowodzenia twierdzeń logicznych, jak np. metoda rezolucji (zob. Szałas 1987). Z badań eksperymentalnych jednak wiadomo, że ograniczenia inteligencji ludzkiej nie pozwalają stosować tych metod podczas myślenia. Nie konstruuje się więc, na gruncie nauki poznawczej, modeli wnioskowania dedukcyjnego opartych na tych metodach. Ograniczenia te nie są jednak brane pod uwagę przez badaczy AI przy budowie systemów sztucznej inteligencji. Nierzadko okazuje się, że systemy te przewyższają zdolnościami człowieka. Słynnym przykładem jest przegrany przez arcymistrza Kasparowa pojedynek szachowy z komputerem Deep Blue. Mimo powyższej różnicy związek obu dziedzin jest bardzo ścisły i pokrywają się one w dużym stopniu. Przepływ myśli następuje w dwie strony – często systemy stworzone początkowo do modelowania procesów poznawczych są rozwijane i uzyskują „własne życie” w nauce o sztucznej inteligencji.

Od psychologii poznawczej odróżnia naukę poznawczą fakt, że psychologia poznawcza bada umysł człowieka, a nauka poznawcza umysł „w ogóle”, czy naturalny czy też sztuczny. Oczywiście, nauki te są także ściśle powiązane, gdyż umysł sztuczny konstruuje się w nauce poznawczej na wzór umysłu ludzkiego. Niektórzy zwolennicy nauki poznawczej (Chalmers 1997) twierdzą, że komputer wyposażony w tak stworzony program nie symuluje myślenia (tak jak program komputerowy symuluje np. przebieg huraganu), ale wręcz myśli.

Teorie w psychologii poznawczej charakteryzują się ponadto mniejszym stopniem formalizmu (Domańska 1991). W nauce poznawczej model zapisany jest w języku formalnym – języku programowania. Psychologia używa zaś diagramów i opisu słownego. Psychologia poznawcza nie ma też tak wyraźnie zakreślonych podstaw teoretycznych, jak nauka poznawcza (Maruszewski 1996, ss. 9-10).

Neurofizjologia dostarcza nauce poznawczej wielu danych eksperymentalnych, a także jest inspiracją dla jednego z dwóch głównych podejść do badania umysłu – koneksjonizmu. W podejściu tym, opisanym w rozdziale 4.3, do przetwarzania informacji używa się systemów o architekturze analogicznej do architektury mózgu. W systemach tych dane przetwarzane są przez dużą liczbę równoległe pracujących elementów – analogonów neuronów. Elementy te, podobnie jak neurony, spięte są za pomocą wielu połączeń w sieć. Neurofizjologia dostarcza także danych o wpływie funkcjonowania mózgu na ograniczenia procesów umysłowych, tłumacząc je m.in. szybkością przewodzenia impulsów przez neurony. Wielu danych o pracy

umysłu dostarczają także badania prowadzone na osobach z uszkodzonym mózgiem. Jako nauka biologiczna neurofizjologia stosuje jednak zupełnie inne metody badawcze i inny jest przedmiot jej badań – nie umysł, lecz mózg.

Lingwistyka bazująca na teorii gramatyki Chomsky'ego zdominowała teoretyczne podstawy symulacji procesów językowych w obrębie nauki poznawczej, charakteryzuje ją także zbliżony stopień formalizmu (Green i in. 1996, rozdz. 7). Jest ważna dla nauki poznawczej także w kontekście języka myśli. Teoria ta, zaproponowana przez Jerry'ego Fodora (Fodor 1976), zakłada istnienie wrodzonego ludziom języka będącego nośnikiem wszystkich procesów poznawczych. Każda myśl, obojętnie czy w postaci zdania czy wyobrażenia przestrzennego, na najbardziej podstawowym poziomie przetwarzana jest, wg Fodora, w postaci zdaniowej. Jest to jakby odpowiednik języka maszynowego w komputerze. W związku z tym ważne jest rozwijanie metod przetwarzania struktur zdaniowych – a metod tych dostarcza lingwistyka.

Filozofia umysłu, jako ostatnia z dziedzin o największym wpływie na naukę poznawczą, daje jej podstawy filozoficzne w postaci funkcjonalizmu. Dostarcza także aparat pojęciowy dotyczący ogólnych zagadnień związanych z umysłem (Salo 1997). Filozofia jest ważna dla nauki poznawczej tym bardziej, że dużo miejsca w tej nauce poświęca się metodologicznej refleksji (Domańska 1991).

Wymienione dyscypliny uzupełniają się w opisie umysłu w ten sposób, że badania interdyscyplinarne prowadzone przy wykorzystaniu osiągnięć tych dziedzin pozwalają analizować zjawiska poznawcze na wszystkich możliwych poziomach opisu. Neurofizjologia bada poziom najniższy – implementacji procesów poznawczych w mózgu. Nauka o sztucznej inteligencji bada umysł na poziomie obliczeniowym – poziomie operacji syntaktycznych. Psychologia dostarcza opisu na poziomie zachowania się podmiotu (poziom najwyższy). Lingwistyka to poziom zarówno syntaktyczny jak i zachowania się podmiotu operującego językiem. Filozofia dostarcza pojęć i refleksji metodologicznej stanowiąc metapoziom opisu (Salo 1997).

Osiągnięcia innych dziedzin nauki są oczywiście wykorzystywane w nauce poznawczej, ale nie wyznaczają jej charakteru tak silnie, jak pięć wspomnianych wcześniej dyscyplin. Czasami bezpośrednio, a czasami na zasadzie analogii, wykorzystywane są badania na gruncie genetyki, teorii ewolucji, zoopsychologii, fizyki kwantowej, pozostałych działów psychologii, socjobiologii, antropologii czy nauk o wychowaniu. Nie można zapomnieć także o kluczowym dla zachowania odpowiedniego stopnia formalizmu aparacie matematyki, logiki czy teorii systemów.

### 3. Metodologia nauki poznawczej

W niniejszym rozdziale opiszę metody, jakimi nauka poznawcza bada procesy poznawcze. Zacznę od przedstawienia pewnej strategii, która choć upraszcza opis zachowania człowieka, to pozwala jednocześnie na użycie ściślejszych metod do tego opisu. Strategia ta, powszechnie przyjęta w nauce poznawczej, zwana jest solipsyzmem metodologicznym. Następnie przedstawię poziomy opis, na jakich poznanie umysłu może się dokonywać. Omówię także założenia nt. pewnych sztywnych, fizycznych ograniczeń umysłu, bez uwzględnienia których nie da się stworzyć poprawnego modelu procesów poznawczych. Wreszcie opiszę same metody: eksperyment i teorię, jako tradycyjne narzędzia nauki oraz

model komputerowy, którego znaczenie na gruncie nauki poznawczej różni się od znaczenia modeli stosowanych w innych naukach.

### 3.1. Solipsyzm metodologiczny

U podstaw metodologii nauki poznawczej leży koncepcja solipsyzmu metodologicznego [*methodological solipsism*] jako strategii badawczej, jaką należy przedsięwziąć, aby naukowe poznanie umysłu było możliwe.

Nazwa ‘solipsyzm’ pochodzi od łacińskich słów ‘*solus*’ – ‘sam’ + ‘*ipso*’ – ‘ja’. Termin ‘solipsyzm metodologiczny’ powstały w latach dwudziestych "przypomniał" Putnam (Putnam 1975, s. 227), a jako strategię badawczą wprowadził do nauki poznawczej Fodor (Fodor 1981, za: Traiger 1991). Strategia ta stanowi zalecenie, aby konstruując model ludzkiej inteligencji założyć (dla potrzeb) istnienie tylko umysłu badanego. To, co jest ważne dla naukowego opisu umysłu, to nie świat zewnętrzny, do którego reprezentacje w badanym umyśle się odnoszą, ale wyłącznie te reprezentacje. Fodor daje przykład programu SHRDLU, który manipuluje bryłami w stworzonym wirtualnie „świecie brył”. SHRDLU formułuje wypowiedzi o bryłach mimo, iż żadne bryły w tym świecie nie istnieją realnie, a tylko programista dostarcza robotowi odpowiednich danych.

Podejście Fodora jest podobne do tezy Sticha, zwanej zasadą autonomii [*principle of autonomy*]. Zasada ta głosi, że stany, które powinny być badane przez psychologów, to wyłącznie stany, które są dodatkowym efektem [*which supervene*] wewnętrznego, fizycznego stanu organizmu. Nie należy więc analizować tego, co dzieje się w środowisku na zewnątrz organizmu. To bowiem, co organizm dostrzegł w otaczającym go świecie, znajduje się jako odpowiednia reprezentacja umysłowa w jego umyśle. To zaś, czego nie zauważył, nie ma wpływu na jego procesy umysłowe (Wilson 1999).

Stanowisko przedstawione w niniejszym podrozdziale ma wielu krytyków. Pamiętać jednak należy, że nauka poznawcza nie zakłada ani tego, że istnieje tylko badany umysł, ani też tego, że dowiemy się wszystkiego o umyśle badając go w oderwaniu od świata zewnętrznego. Strategia solipsyzmu metodologicznego jest po prostu użyteczna (choć upraszczająca) w badaniu przez dopiero co powstałą naukę niezwykle skomplikowanego zjawiska, wobec którego nie dysponujemy metodą bezpośredniej obserwacji. Brak jest także teorii, która zadowalająco wyjaśniałaby, jak reprezentacje obecne w umyśle odnoszą się do świata na zewnątrz.

### 3.2. Poziomy opisu umysłu

Naukowcy poznawczy rozważając umysł jako system przetwarzający informację na sposób obliczeniowy muszą określić ogólne zasady działania właściwe każdemu egzemplarzowi klasy – jego architekturę funkcjonalną (w opozycji do architektury anatomicznej<sup>7</sup>). Architektura ta zwana jest w nauce poznawczej, jako że odnosi się do systemów poznawczych, architekturą poznawczą.

Przyjmuje się, że zarówno komputery, jak i umysły ludzkie posiadają trzy odrębne poziomy organizacje, wyznaczające ich architekturę poznawczą (Pylyshyn 1989, s. 57):

<sup>7</sup> Lepiej znaczenie oddałby termin ‘strukturalnej’. Przep. M. Kasperski.

1. Poziom semantyczny (zwany poziomem wiedzy); na tym poziomie szuka się wyjaśnienia, dlaczego ludzie i odpowiednio zaprogramowane komputery zachowują się tak, a nie inaczej, w terminach ich przekonań, celów i wiedzy, pokazując, że owe cele i przekonania połączone są w sposób racjonalny i posiadający pewne znaczenie.
2. Poziom syntaktyczny (zwany poziomem symboli); wiedza, cele i przekonania systemu na tym poziomie są przedstawione w postaci symboli i syntaktycznych na nich operacji.
3. Poziom fizyczny (biologiczny); system poznawczy, aby działał, musi być realizowany w jakiejś fizycznej formie, na tym poziomie opisuje się strukturę i funkcjonowanie tego fizycznego lub biologicznego substratu.

Według amerykańskiego filozofa Daniela Dennetta, istnienie owych trzech poziomów opisu pozwala nam określić zarówno umysły ludzkie, jak i komputery jako systemy intencjonalne. Tylko w stosunku do systemów intencjonalnych możemy przyjąć trzy poniższe podejścia [*stances*] badawcze. Zetknąwszy się z dowolnym obiektem można próbować zrozumieć jego funkcjonowanie analizując właściwą mu konstrukcję fizyczną. Jest to podejście fizyczne [*physical stance*]. Jest ono o tyle niezawodne, o ile niezawodne są prawa fizyki (pominąwszy efekty kwantowe), lecz jednocześnie niezmiernie skomplikowane. Możemy więc próbować zrozumieć działanie owego obiektu pomijając zagmatwane szczegóły jego budowy i badając go jak gdyby był zaprojektowany do spełniania jakiejś funkcji. Przyjmujemy wtedy podejście projektowe [*design stance*]. Jest to częsta strategia w stosunku do urządzeń stworzonych przez człowieka. Mało kto wie przecież jak zbudowany jest komputer, ale jak działa – wie prawie każdy. W przypadku umysłu człowieka wystarczy przyjąć, że owym projektantem była natura – choć działająca bez celu, to tworząca dzięki ewolucji obiekty niezwykle skomplikowane i precyzyjne. Wg Dennetta metoda ta jest efektywna i niezawodna, jeśli tylko znamy dokładnie projekt urządzenia i urządzenie to nie zacznie funkcjonować niepoprawnie. Wreszcie, możemy założyć, że badany obiekt zachowuje się racjonalnie, ma własne cele i przekonania, i będzie się zachowywał tak, aby dążyć do owych celów w oparciu o posiadaną wiedzę. Jest to istotą podejścia intencjonalnego [*intentional stance*]. Pozwala ono przewidywać zachowanie się obiektu, który jest na tyle złożony, że dokładnie nie można poznać jego fizycznej konstytucji i właściwych mu zasobów funkcjonowania. Systemy intencjonalne to systemy, których zachowanie można przewidzieć analizując ich intencje: cele, przekonania, zamiary itp. (Dennett 1987, ss. 16-25).

Dennett i jego zwolennicy wierzą, że w przypadku systemów intencjonalnych powyższe poziomy są powiązane przyczynowo: cel czy przekonanie pojawiające się na poziomie semantycznym (intencjonalnym) powoduje powstanie określonej struktury symbolicznej na poziomie syntaktycznym (funkcjonalnym), realizowanej w specyficzny dla siebie sposób fizyczny. Logiczna forma przekonań i celów będzie odzwierciedlana w strukturze fizycznego substratu. Wg Dennetta opis zachowania się człowieka jest zatem sprowadzalny do terminów fizycznych, ale prowadzi to do eksplozji kombinatorycznej – wymaga tak wielu obliczeń, że żaden system nie jest w stanie ich wykonać w rozsądnym czasie. Wygodniej zatem wyjaśniać zachowanie człowieka (i komputera) na poziomie syntaktycznym (tamże, ss. 35-36). Jest to przekonanie dosyć rozpowszechnione w nauce poznawczej, choć kwestionowane przez niektórych naukowców poznawczych (m.in. Fodora i Pylyshyna). Uważają oni, że możliwe jest opisanie zachowania człowieka jedynie na poziomie syntaktycznym (podejście projektowe), gdyż nie istnieje, wg nich, jednoznaczna relacja między stanami fizycznymi mózgu a pracą umysłu (zob. Lyons 1995, ss. X-XI).

Aby wyjaśnić wzajemną odpowiedniość poziomów opisu można posłużyć się następującym przykładem: Załóżmy, że prowadzimy grę w szachy z szachowym komputerem. Chcąc wygrać partię musimy stosować się do reguł szachowych oraz dokonywać najlepszych posunięć, przewidując jednocześnie zachowanie przeciwnika. Traktujemy więc program szachowy jako system intencjonalny. Bierzymy pod uwagę jego cel (wygrać z człowiekiem) i jego wiedzę (m.in. reguły gry i zapamiętane wcześniejsze partie). Wiemy, że będzie się zachowywał tak, aby zrealizować swój cel – racjonalnym jest oczekiwanie, że zbije naszą figurę, jeśli będzie miał ku temu okazję<sup>8</sup>. Znając jednak program komputera moglibyśmy analizować jego posunięcia czysto syntaktycznie. Znając formalne reguły wyboru posunięć przez komputer możemy dokładnie przewidywać jego grę. Jest to skomplikowane, ale możliwe. Możemy także analizować zachowanie komputera nie znając jego programu. Dokonać możemy tego na najniższym poziomie opisu – poziomie języka maszynowego, odpowiadającego rozkładowi ładunków elektrycznych w samym komputerze. Zwiększa to jednak drastycznie zakres koniecznych obliczeń, nie dostarczając dodatkowej wiedzy o sposobie gry komputera. Już dzięki analizie programu dowiedzieliśmy się bowiem co robi komputer (zakładając, że działa on poprawnie). Dzięki podejściu fizycznemu dowiedzieliśmy się tylko jakie procesy fizyczne realizują zachowanie. Dennett wierzy, że w ten sam sposób możemy opisywać zachowanie. Wiedza o pracy mózgu człowieka jest zatem niepotrzebna do wyjaśnienia jego normalnego zachowania. Pogląd ten łączy znakomitą większość badaczy poznawczych. Wiedza ta zyskuje natomiast znaczenie w przypadku uszkodzeń mózgu i jego wadliwego funkcjonowania – podobnie jak sama analiza programu komputera nie wystarczy, gdy nastąpi awaria procesora.

W analizie systemów koneksjonistycznych postuluje się także istnienie dodatkowego poziomu pomiędzy poziomem fizycznym, a poziomem symbolicznym (Simon, Kaplan 1989, s. 8). Ów poziom jest już sposobem opisu w terminach formalnych, a nie fizycznych, lecz nie operuje symbolami, a stanem aktywacji sieci prostych elementów. Zwany jest poziomem subsymbolicznym. Prowadzone są badania mające na celu wyjaśnienie zależności pomiędzy poziomami symbolicznym i subsymbolicznym poprzez budowę systemów hybrydowych połączonych niefizycznie [*non-physically hybrid systems*] (zob. rozdz. 4.5).

Naukę poznawczą interesuje poziom symboli (lub subsymboli). Lecz przy opisie systemów inteligentnych na tym poziomie pojawia się problem: Skąd wiadomo, że struktura procesów u dwóch zachowujących się tak samo obiektów (np. komputera i człowieka) jest także taka sama? Może się przecież zdarzyć, że choć wynik danego procesu jest identyczny u obydwu (przy identycznych danych wejściowych), to przebiegają one w zupełnie innych sposób. Mówimy wtedy o słabej równoważności procesów. Aby symulować procesy poznawcze człowieka w komputerze należy zagwarantować więc choć w przybliżeniu ich silną równoważność. Jest ona zapewniona, gdy obydwie systemy realizują ten sam algorytm, choć najczęściej w różnych implementacjach (Pylyshyn 1989, s. 171). Istnieją różne metody ustalania silnej równoważności procesów. Polegają one na uzgadnianiu procesów przebiegających w komputerze (do których struktury mamy dostęp) z wynikami badań psychologicznych, a także coraz częściej neurofizjologicznych, które tylko pośrednio tłumaczą strukturę procesów w niedostępnym bezpośredniemu badaniu (poza introspekcją) umyśle. Metody te przedstawię w następnym podrozdziale.

<sup>8</sup> Na marginesie dodam, że w taki sposób zachowywał się Garry Kasparow, grając mecz z Deeper Blue. Maszyna nie skorzystała ze zręcznie podsuniętej przez Kasparowa okazji zbiccia figury, co między innymi w ostateczności przyczyniło się do przegranej arcymistrza. Przyp. M. Kasperski.

Podobnie, jak jest problemem uzgodnienie wzajemnej relacji między poziomem semantycznym a syntaktycznym, tak i zbadania wymaga wpływ implementacji procesów poznawczych w konkretny system fizyczny (mózg lub komputer) na przebieg tychże procesów. Widza i cele systemu poznawczego stanowią struktury złożone z pewnych podstawowych elementów. Ponieważ postać tych struktur zależy od modyfikowalnej wiedzy i celów, są one „penetrowalne poznawczo”. Zmieniając wiedzę systemu i obserwując zmianę jego zachowania, możemy próbować określić te struktury. Jednak podstawowe elementy, z których składają się reprezentacje i najbardziej podstawowe operacje systemu poznawczego takiej penetracji się już nie poddają. Kształt najmniejszych „cegiełek” systemu nie zależy od zawartości poziomu semantycznego, dopiero ich struktury ulegają takiej zależności. Na kształt podstawowych elementów systemu ma jednak wpływ poziom fizyczny, odzwierciedla on ograniczenia i zasoby substratu fizycznego. A ponieważ owe „cegiełki” wyznaczają możliwe operacje na poziomie syntaktycznym, zależy od nich, zatem także system przekonań i celów (tamże, ss. 72-81).

Informacje na temat architektury funkcjonalnej systemu poznawczego należy więc czerpać z wiedzy neurofizjologicznej. Np. fakt, że większość procesów, tj. percepcja, przypominanie czy rozumienie zdań, zajmuje nie więcej niż sekundę, a czas reakcji neuronu mierzy się w milisekundach, nakłada na system ograniczenie zwane „ograniczeniem programu 100 kroków”. System ma od dyspozycji tylko ok. 100 faz, etapów, w ciągu których musi zakończyć dany proces (Rumelhart 1989, s. 135). Innym sposobem określenia podstawowych jednostek ludzkiego poznania jest analizowanie w ramach psychologii tego, co niezmiennie w zachowaniu niezależnie od wiedzy, celów i przekonań (Pylyshyn 1989, s. 81). Przykładem może być opisywana w rozdziale pierwszym hipoteza nt. pojemności pamięci krótkotrwałej. Pojemność ta nie zmienia się w zależności od tego, co jest pamiętane. Czy są to cyfry numeru telefonu komornika, czy też telefonu do najpiękniejszej nawet dziewczyny w mieście, możemy tych cyfr przechowywać w pamięci krótkotrwałej zaledwie kilka.

Architektura funkcjonalna wyznacza możliwości obliczeniowe systemu poznawczego. Interesująca jest więc nie tylko zależność: jakie ograniczenia nakłada określona biologiczna i fizyczna budowa mózgu lub komputera na przebieg procesów poznawczych (a więc realizacja jakich procesów możliwa jest w danej architekturze), ale i zależność odwrotna: jaka powinna być struktura fizyczna obiektu, w którym chcemy realizować (implementować) dany abstrakcyjny system poznawczy (a więc w jakich architekturach możliwa jest realizacja danego procesu). Jedną z możliwych odpowiedzi jest zaproponowany przez Chalmersa (Chalmers 1997) następujący warunek wystarczający do tego, aby system fizyczny implementował zadany proces obliczeniowy: struktura fizyczna musi odzwierciedlać formalną (logiczną) strukturę procesu obliczeniowego. Musi zatem istnieć pogrupowanie pewnej liczby fizycznych stanów w pewien zbiór  $F_n$ , takie, że danemu formalnemu (logicznemu) stanowi  $L_n$  jest ów zbiór jednoznacznie przyporządkowany w taki sposób, że istnieją odpowiadające sobie relacje, z których jeśli jedna (formalna) przyporządkowuje pewnemu stanowi formalnemu  $L_n$  inny stan formalny  $L_k$ , to druga przyporządkowuje dowolnemu stanowi fizycznemu ze zbioru  $F_n$  (odpowiadającemu stanowi  $L_n$ ) dowolny stan fizyczny ze zbioru  $F_k$  (odpowiadającego stanowi  $L_k$ ).

### 3.3. Modułowość

Następnym zagadnieniem dotyczącym architektury poznawczej jest problem modułowości umysłu. Czy system poznawczy funkcjonuje jako jedna struktura, czy też zawiera podsystemy realizujące właściwe im zadania niezależnie od tego, co robi reszta systemu? Dość

rozpowszechnioną odpowiedzią na powyższe pytanie jest koncepcja modułowości umysłu J. Fodora, wyodrębniająca trzy rodzaje podsystemów:

1. przekaźniki [*transducers*] – organy zmysłowe odbierające sygnały ze świata zewnętrznego,
2. systemy wejścia – dostarczają przetworzone informacje z organów zmysłowych do centralnych,
3. systemy centralne – odpowiadają za procesy umysłowe wysokiego poziomu.

Systemy modułowe wejścia m.in. są:

- szczegółowego przeznaczenia [*domain-specific*] i realizują specyficzne dla nich procesy poznawcze,
- enkapsulują informację niezbędną do wykonywania ich funkcji,
- działają niezależnie od naszej woli,
- dzięki powyższym trzem cechom działają szybko.

Istnieje jednak klasa procesów poznawczych, które wydają się nie być zmodularyzowane – są to tzw. procesy centralne. Ustalanie czyichś sądów czy podejmowanie decyzji (szczególnie skomplikowanych) wymaga brania pod uwagę informacji z różnych modułów i nie podlega logicznym zasadom wynikania [*nondemonstrative inference*], tzn. konkluzja nie wynika w jakiś formalny sposób z przesłanek, lecz jest nimi niewątpliwie powodowana. Weryfikacja hipotez uzyskanych w ten sposób przypomina weryfikację hipotez naukowych, podobnie jest z ich wymyślaniem, lecz trudno stwierdzić jak ten proces przebiega. Nie jest on bowiem enkapsulowany w konkretnym module, lecz wymaga aktywności wielu podsystemów poznawczych i koordynacji ich przez system centralny. Występowanie systemów modułowych i centralnych ma uzasadnienie ewolucyjne. Procesy modułowe (np. percepcja) są szybkie i działają niezależnie od naszych przekonań, pozwalają nam w optymalny sposób odbierać nawet najbardziej niespodziewane informacje z otoczenia (np. dot. zagrożenia). Ilość tych informacji z czasem przekroczyła jednak możliwości obliczeniowe mózgu, powstały więc w trakcie ewolucji systemy integrujące i selekcjonujące napływające dane w sposób wolniejszy, lecz pewniejszy. Niektórzy naukowcy poznawczy nie zgadzają się z argumentacją Fodora. Twierdzą, że systemy centralne są też zmodularyzowane, lecz mają bogate połączenia z innymi systemami (Green i in. 1996, ss. 61-77).

Ściśle związany z zagadnieniem modułowości jest problem języka. Czy procesy językowe są tylko pewnym modulem odpowiadającym za komunikację językową o zupełnie innej organizacji niż np. procesy percepcji, motoryki, podejmowania decyzji itp.? Czy też procesy językowe nakładają się na wszystkie inne procesy, stanowiąc pewne medium, podstawę, wewnętrzne narzędzie umysłu, za pomocą którego przebiega myślenie? O ile na gruncie nauki poznawczej powszechnie przyjmuje się modułowość języka, to istnieją różnice w usytuowaniu owego modułu wobec innych procesów poznawczych. Twierdzi się bądź, że język jest modulem niezależnym od poznania, bądź też, że jest modulem nadbudowanym na jego szczycie (Pinker 1989, s. 360).

Podane tutaj zagadnienia dotyczą każdej postulowanej postaci systemu poznawczego. Są one jednak zbyt ogólne, aby bez dodatkowych założeń proponować konkretny model ludzkiego poznania. Dodatkowe założenia na temat podstawowych operacji obliczeniowych i postaci podstawowych jednostek reprezentacji skutkują w powstawaniu różnych modeli umysłu.

Cztery najważniejsze rodziny takich systemów zostaną zaprezentowane szczegółowo w rozdziale 4.

### 3.4. Metody badawcze

Aby opisać i wyjaśnić pewne zjawisko należy stworzyć jego teorię: usystematyzowany system praw i definicji. Teoria opierając się na prawach, czyli sądach ogólnych, ma zarazem oparcie w faktach, czyli sądach jednostkowych. Prawa systematyzują sąd jednostkowy, ale przede wszystkim pozwalają na ich wyjaśnienie i przewidywanie (Krajewski 1998, cz. I).

W nauce poznawczej prawami, jakie należy odkryć, są prawa dotyczące procesu poznania u człowieka: zależności pomiędzy bodźcami w świecie zewnętrznym, reprezentacją owego świata w umyśle człowieka i jego zachowaniem. Do ustalenia faktów dotyczących zachowania się człowieka w różnych warunkach oraz do odkrycia praw, które fakty te wyjaśniają, a wreszcie w celu weryfikacji i falsyfikacji tychże praw nauka poznawcza posiada szereg metod poznawczych. Wśród nich można wyróżnić metody przejęte wraz z wiedzą pochodzącą z dziedzin składowych (a głównie z psychologii) oraz metody swoiste nauce poznawczej.

Ponieważ powszechnie przyjmuje się w nauce poznawczej modułową naturę umysłu, tworzy się w związku z tym teorie dotyczące wybranego subsystemu poznawczego. Pomijając zresztą aspekt merytoryczny ma to sens praktyczny. Rozłożenie niezwykle skomplikowanego problemu (a badanie umysłu z pewnością takim problemem jest) na podproblemy skutkuje wzrostem prostoty i w przypadku umysłu jest przynajmniej na razie konieczne.

Próba zrozumienia procesu poznawczego polega na stosowaniu metod układających się w pewien schemat. Poprzez zbieranie faktów, konstrukcję i weryfikowanie hipotez tworzona jest teoria. Stosowane są w tej fazie tradycyjne eksperymenty (głównie psychologiczne, ale także lingwistyczne i inne) oraz znane metody weryfikacji. W tej fazie odznacza się zarazem duży wpływ aparatu pojęciowego nauki o sztucznej inteligencji – badane są aspekty informacyjne/obliczeniowe zjawiska. Określony zbiór danych może być jednak opisywany poprawnie przez różne teorie. Aby upewnić się, że konkretna teoria jest najlepsza, przechodzi się do drugiej fazy. Polega ona na budowie na bazie wypracowanej teorii modelu obliczeniowego, implementowanego w postaci programu w komputerze. Ma to ogromne znaczenie dla procesu badawczego: o ile teoria nie jest w pełni określona i nie precyzuje wszystkich swoich elementów, to fizycznie działający model musi być jednoznaczny, niesprzeczny i spójny. O ile procesy, które opisuje teoria nie są dla naukowca do końca dostępne, o tyle procesy wewnątrz modelu dają się dobrze podczas symulacji obserwować. Stworzenie modelu świadczy, że teoria jasno wyjaśnia zależności między dwoma pojęciami i że jej postulaty są fizycznie realizowalne.

Konkretyzacja teorii pociąga za sobą także szereg zagrożeń. Zmuszony do precyzji programista może (i zazwyczaj tak czyni) wprowadzić do modelu zależności, których teoria nie postulowała. Krytycy nauki poznawczej podnoszą wtedy zarzut, że to nie program jest inteligentny, ale wyłącznie programista. Dodatkowo, model realizowany w konkretnym języku programowania odzwierciedla ograniczenia tegoż języka. Ograniczenia te mogą się znacznie różnić od ograniczeń obliczeniowych ludzkiego umysłu. Wymienione zagrożenia to wady nieusuwalne. Można je jedynie częściowo ograniczyć, lecz mimo to praktyczna i metodologiczna korzyść ze stosowania symulacji jest ogromna. Zresztą, jak zauważają filozofowie nauki (Krajewski 1998, s. 116), problemy idealizacji – pomijania



przeszkadzających warunków ubocznych – trapią wszystkie nauki, nawet współczesną fizykę. Nie oznacza to jednak, że naukowcy powinni zarzucić jej uprawianie.

Dysponując działającym modelem, który zachowuje się zgodnie z teorią: odzwierciedla i przewiduje zachowanie się człowieka w opisywanej przez teorię dziedzinie, stajemy przed problemem ustalenia stopnia równoważności procesów naturalnych i modelowanych. Poszukujemy ich silnej równoważności. Metody jej ustalania są metodami swoistymi nauce poznawczej. W innych naukach między zjawiskiem a jego symulacją w komputerze istnieje ogromna różnica, chodzi o uzyskanie właściwych wyników. Żadnego astronoma nie wierzysz, że w komputerze obraca się miniwszechświat, ani żaden meteorolog nie sądzi, iż w komputerze wieje minihuragan. W nauce poznawczej sprawa przedstawia się zupełnie inaczej. Znacząca większość naukowców poznawczych (por. Chalmers 1997) rozpatruje umysł jako pewną maszynę, o dużym stopniu podobieństwa funkcjonalnego do komputera i oba obiekty rozpatruje jako zdolne do myślenia.

Po ustaleniu stopnia równoważności procesów koło się zamyka: czynione są nowe obserwacje i eksperymenty (aby uzyskać więcej danych), modyfikowana jest teoria (aby uwzględnić nowe dane i wnioski z badania starego modelu) i konstruowany jest nowy model. Oczywiście kolejne cykle owego koła następują przy założeniu, że osiągany jest jakiś postęp. Dokonujemy zatem przeglądu metod stosowanych w opisanych powyżej fazach.

Metody empiryczne stosowane w pierwszym etapie dzielą się na metody obserwacyjne i eksperymentalne. Metody obserwacyjne są proste w stosowaniu, ale obciążone wieloma wadami. Dwie najpopularniejsze to: obserwacja w naturalnym środowisku, gdzie trudno jednak oddzielić wpływ dodatkowych czynników oraz analiza korelacji zjawisk, nie mówiąca niestety wiele o zależności przyczynowo-skutkowej. Wśród metod eksperymentalnych istnieje wiele ich rodzajów, zależnie od badanej domeny i inwencji naukowca. Wspólną ich cechą jest to, że służą testowaniu konkretnej, sprecyzowanej hipotezy i czynią to w specyficzny sposób. W eksperymencie psychologicznym naukowiec manipuluje cechą zwaną zmienną niezależną (bodźcem) i bada jej wpływ na zachowanie się obiektu, jego reakcję rozważaną jako zmienną zależną. Aby eksperyment był wiarygodny zmienna zależna powinna mieć charakter ilościowy (choć czasami obserwacje jakościowe są nieuniknione), efekty przyczynowo-skutkowe wyraźnie oddzielone (np. poprzez losowy dobór osób do dwóch grup – badanej i kontrolnej, nie informowanie badanych o celach badania itp.) a eksperyment musi dać się powtórzyć (Bower i Clapper 1989, ss. 246-248).

Dane empiryczne mogą pochodzić także z wielu innych dziedzin, zarówno z nauk przyrodniczych jak i społecznych. Ważnym ich źródłem jest neurofizjologia, dostarczająca danych na temat mózgu i wpływu jego funkcjonowania na przebieg procesów poznawczych. Szczególnie ważne są tu dane dotyczące zachowania się osób z uszkodzonym mózgiem (świadczące m.in. o analizowanej w poprzednim rozdziale modułowości systemu poznawczego), dane z elektroencefalogramów, technik magnetycznego rezonansu jądrowego, tomografii pozytronowej, badań histopatologicznych czy nawet rejestracji aktywności pojedynczych neuronów (zob. Sejnowski i Churchland 1989, rozdz. 8.2).

Antropologia, psychologia społeczna i socjologia dostarczają danych o wpływie środowiska społecznego na zachowanie, a także o przebiegu procesów umysłowych w różnych kulturach. Przykładem mogą być badania wnioskowania sylogistycznego (Scribner 1977), które pokazują, że w pewnych prymitywnych społecznościach (farmerskie plemię Kpelle w Liberii) proces wnioskowania nie prowadzi do abstrakcyjnej analizy przesłanek, ale do odniesienia ich

do rzeczywistości. Jeśli przesłanka nie opisuje sytuacji znanej członkowi Kpelle, nie potrafi on podać rozwiązania nawet najprostszego sylogizmu. Pozostając w kręgu kultury europejskiej byłibyśmy nieświadomi takich zjawisk – naszym teoriom brak byłoby uniwersalności.

Wreszcie, dane pochodzą z tak wydawałoby się odległej nauce poznawczej dziedziny, jak historia nauki. Badając przebieg odkrycia naukowego, zapiski odkrywcy i jego późniejsze zeznania, stawia się hipotezy na temat rozwiązywania pewnej klasy problemów – procesów poznawczych prowadzących do odkryć naukowych. Zwane jest to analizą zawartości [*content analysis*]. Pewną wersją tej metody, towarzyszącą jednak dużo bardziej standardowym odkryciom, jest tzw. analiza protokołu [*protocol analysis*]. Polega na tym, że w trakcie rozwiązywania zadanego problemu badany mówi na głos kolejne myśli. Są one później analizowane w celu oddzielenia danych nieistotnych, zasugerowanych lub przeinaczonych, metoda ta jest bowiem wersją introspekcji i dziedziczy jej ograniczenia (Simon i Kaplan 1989, ss. 21-27).

Do budowy modelu wykorzystywana jest wiedza pochodząca z nauki o sztucznej inteligencji, przede wszystkim na temat inteligentnych systemów obliczeniowych i możliwych rodzajów reprezentacji. Ponieważ praca niniejsza traktuje naukę poznawczą właśnie pod kątem jej związków z nauką o sztucznej inteligencji, zagadnienia te zostały omówione w rozdziałach czwartym i piątym.

Oprócz wiedzy na temat budowy konkretnego modelu ważna jest też znajomość historii konstruowania sztucznych systemów poznawczych, struktury ich zbioru, osiągniętych sukcesów i niepowodzeń. Temu celowi służyć ma metoda zwana metaanalizą. Polega ona na użyciu metod statystycznych do badania struktury i tendencji w studiach poznawczych. M.in. jedna z pierwszych metaanaliz, przeprowadzona na 23 modelach opublikowanych w latach 1980-86 w czasopiśmie "Cognitive Science" pokazała przewagę symulacji szeregowych (14) nad równoległymi (7, 2-nieokreślone), symulacji rozwiązywania problemów (6) i procesów językowych (4) nad pozostałymi dziedzinami. Wskazano także na fakt, że tylko nieliczne modele (2) umożliwiały dokładne i trafne przewidywanie wyników eksperymentów na ludziach (Simon i Kaplan 1989, ss. 31-33).

Kluczowe dla nauki poznawczej są badania ustalające stopień równoważności procesów poznawczych zachodzących w umyśle i w programie. Spośród pewnej liczby technik oceny stopnia owej zależności jedną z najistotniejszych jest wykorzystywanie czasowych parametrów poznawczych. Metoda ta bada wzajemne stosunki pomiędzy czasem wykonywania różnych zadań: jeśli zarówno człowiek jak i maszyna najdłużej wykonują zadanie Z, krócej zadanie Y, a najkrócej zadanie X, można przypuszczać, że obydwa systemy wykonują te zadania podobnie (Domańska, 1991). Prosty przykładem może być wybór spośród dwóch teorii dodawania małych liczb naturalnych przez dzieci. Pierwsza teoria stwierdzała, że dzieci dodając dwie liczby wykonują operację sumowania. Druga – że dzieci dodają zwiększając większy składnik o jeden tyle razy, ile wynosi mniejszy składnik. Choć teorie prognozują ten sam czas wykonywania operacji „4+1”, tylko druga przewiduje zaobserwowany fakt, że operacja „7+1” zabiera mniej czasu niż „7+3”. Pierwsza operacja wykonywana jest w jednej fazie (7+1), druga wymaga trzech faz (7+1+1+1) (Green i in. 1996, s. 19).

Opisane powyżej metody nie doprowadziły jak dotychczas do rozstrzygnięcia ani tego, czy umysł ma naturę obliczeniową, ani tym bardziej tego, jaka owa natura jest. Mimo to stanowią

potężne i obiektywne narzędzie w rękach naukowców poznawczych, łącząc w sobie najlepsze cechy metod pochodzących z różnych dziedzin. Pozwalają przede wszystkim na tworzenie modeli procesów poznawczych, które są precyzyjne, niesprzeczne, spójne oraz w dużym stopniu pozbawione niedomówień i nieokreślonych pojęć. Co więcej, systemy symulujące umysł pozwalają na przewidywanie zachowania się ludzi – mają moc predykcji, co jest chyba najważniejszą cechą praw naukowych. Nie jest bowiem sztuką tylko wyjaśnienie pewnego zjawiska na podstawie zebranych danych, szczególnie gdy dotyczy to tak trudno poddającego się obserwacji umysłu i możliwych wyjaśnień może być wiele. Sztuką jest wygenerowanie danych o przyszłych stanach opisywanego zjawiska i ich pozytywne zweryfikowanie. Wiele spośród teorii psychologicznych (np. psychoanaliza) stara się w sposób koherentny tłumaczyć zachowanie się człowieka, nie potrafi go jednak skutecznie przewidzieć. Jak się podkreśla (Krajewski 1998, s. 51) bez możliwości przewidywania za pomocą praw nauki, niemożliwe jest jednak ich sprawdzenie.

#### 4. Obliczalność i systemy obliczeniowe w nauce poznawczej

Pojęcie obliczalności [*computability*] (obliczenia [*computation*]) jest fundamentalnym pojęciem nauki poznawczej, jest właściwie częścią definicji tej nauki. Nie należy go jednak analizować w oderwaniu od drugiego kluczowego dla nauki poznawczej pojęcia – reprezentacji. Każde obliczanie polega bowiem na operacjach na reprezentacjach (symbolach, znakach, aktywacji neuronu itp.). Pojęć dotyczących niektórych reprezentacji (np. sieci semantycznych) użyto więc w niniejszym rozdziale mimo, iż reprezentacje te opisane zostaną dopiero w rozdziale następnym.

Centralną rolę obliczalności dla nauki poznawczej podkreślają jej dwie tezy: teza o obliczeniowej wystarczalności [*computational sufficiency*] mówiąca, że odpowiednia struktura obliczeniowa systemu wystarczy, aby posiadał on umysł (a co za tym idzie całą gamę umysłowych atrybutów) oraz teza o obliczeniowej wyjaśnialności [*computational explanation*], głosząca, że procesy obliczeniowe zapewniają ogólne narzędzie wystarczające do badania procesów poznawczych i zachowania (Chalmers 1997).

##### 4.1. Definicja obliczalności i przykłady systemów obliczeniowych

Definicja obliczalności powstała na gruncie matematyki i logiki podczas badań prowadzonych w poszukiwaniu odpowiedzi na pytanie zadane przez słynnego matematyka niemieckiego Dawida Hilberta: czy może istnieć pewna ogólna, sformalizowana procedura (algorytm) rozwiązywania wszystkich problemów matematycznych (czyli dowodzenia twierdzeń) należących do pewnej dobrze zdefiniowanej klasy? Pierwszą, negatywną zresztą, odpowiedź sformułował Kurt Gödel: wnioskiem z jego dowodu jest to, iż obok problemów, które można rozwiązać metodą sformalizowaną (algorytmem), czyli problemów obliczalnych, z konieczności muszą istnieć problemy, dla których jest to niemożliwe. Wkrótce podano wiele przykładów takich problemów. Brakowało jednak w czasach Gödela zdefiniowanego pojęcia mechanicznej procedury, czyli algorytmu (Penrose 1996, s. 50).

Pojęcie powyższe zdefiniował w 1936 roku angielski matematyk Alan Turing. Zaproponował, jako najogólniejszy model systemu obliczeniowego, wyidealizowaną maszynę matematyczną nazwaną później od jego nazwiska ‘maszyną Turinga’. Maszyna ta ma skończoną liczbę stanów wewnętrznych (jako, że na każdy algorytm składa się skończona liczba operacji) oraz pamięć o nieskończonej pojemności (gdyż algorytm jest uniwersalny w tym sensie, że nie

zakłada maksymalnej wielkości danych wejściowych). Maszynę tę Turing wyobrażał sobie jako taśmę z liniowo uporządkowanymi dwoma rodzajami symboli (np. „0” i „1”). Algorytm, jaki maszyna ma wykonać, jest zapisany w postaci tablicy, która każdej parze «stan, symbol wejściowy» przyporządkowuje trójkę «stan, symbol wyjściowy, przesunięcie taśmy». Przesunięcie taśmy oznacza miejsce, z którego należy pobrać następny symbol. Maszyna rozwiązuje dany problem przeprowadzając ciąg operacji. Każda z operacji składa się z następującej sekwencji: (1) maszyna odczytuje symbol z taśmy; (2) zależnie od stanu, w którym obecnie się znajduje oraz od odczytu symbolu przechodzi w inny stan; (3) zapisuje na taśmie określony w tablicy symbol (taśma to zarówno pamięć danych wejściowych jak i danych generowanych przez algorytm); (4) przenosi głowicę w określone przez tablicę miejsce. Wyróżnionym stanem maszyny jest „koniec pracy”, w którym kończy ona wykonywanie algorytmu. Dopiero wtedy można odczytać z taśmy wynik jej działania. O ile konkretna maszyna Turinga realizuje konkretny algorytm, to można też w oparciu o powyższy przepis zaprojektować ‘uniwersalną maszynę Turinga’. Maszyna ta jako dane wejściowe otrzymuje kompletny opis konkretnej maszyny Turinga i w oparciu o ten opis wykonuje dokładnie ten sam algorytm, który wykonałaby maszyna konkretna. Daną funkcję możemy nazwać obliczalną, mechaniczną czy algorytmiczną, jeśli da się dla jej realizacji zaprojektować odpowiednią maszynę Turinga. Zakres pojęcia problemów obliczalnych wyznacza więc zbiór problemów dających się rozwiązać za pomocą uniwersalnej maszyny Turinga (tamże, ss. 51-57).

Równocześnie z Turingiem powstała inna definicja funkcji obliczalnej. Odwoływała się nie do analogii maszyny, ale do systemu (języka) formalnego. Jej twórcą był amerykański logik Alonzo Church (przy pomocy Stephena Kleene’a). Stworzył on abstrakcyjny rachunek zwany rachunkiem lambda ( $\lambda$ ), który operuje klasą obiektów: funkcjami (oznaczanymi:  $a, \dots, f, \dots, z, a', \dots, z', a'', \dots$ ). Argumentem dowolnej funkcji jest także funkcja, w wyniku działania funkcją na funkcję otrzymujemy ponownie funkcję. Jedyną operacją tego rachunku jest operacja abstrahowania ( $\lambda$ ) umożliwiająca podstawienie dowolnej funkcji do pewnego wzoru określającego operację. Operacja zdefiniowana np.  $\lambda x.[fx]$ , gdzie  $x$  to zmienna, za którą podstawiamy dowolną funkcję, a  $f$  to określona funkcja, po podstawieniu za  $x$  funkcji  $a$  daje:  $[(\lambda [fx]) a] = fa$ , czyli  $\lambda x.[fx]$  to po prostu funkcja  $f$ . Okazuje się, co udowodnili niezależnie Church i Turing, że w tak prostym języku możemy wyrazić każdy algorytm, dla jakiego istnieje pewna maszyna Turinga. Każdą maszynę Turinga możemy przedstawić w postaci pewnej operacji  $\lambda$ . Na rachunku  $\lambda$  opiera się zaś w istotny sposób (zawiera ten rachunek) język LISP, w którym pisane są niektóre spośród systemów sztucznej inteligencji (Tamże, ss. 85-89).

Istnieje wiele innych równoważnych definicji funkcji obliczalnych, czyli takich, dla których istnieje efektywna (skończona) metoda obliczania wartości dla dowolnego ciągu argumentów. M.in. w jednej z definicji wykorzystuje się rachunek funkcji rekursywnych (rekurencyjnych), w którym w oparciu o trzy podstawowe funkcje: stałą ( $Z [x] = 0$ ), następnika ( $S [x] = x + 1$ ) i tożsamościową ( $I [x] = x$ ) i za pomocą trzech operacji tworzy się z funkcji podstawowych funkcje złożone. Każda funkcja, którą da się w ten sposób uzyskać jest funkcją obliczalną (inaczej rekursywną lub rekurencyjną) (Borkowski 1991, ss. 313-314).

Ostatecznie precyzuje pojęcie obliczalności teza Churcha-Turinga, mówiąca, że to, co możemy nazwać procedurą algorytmiczną, to taka procedura, dla której da się zaprojektować odpowiednią maszynę Turinga (jest ona więc najsilniejszym systemem obliczeniowym) lub pojęcie tej maszynie równoważne w innym równoważnym systemie formalnym (np. operację  $\lambda$  w rachunku  $\lambda$ ). Cecha obliczalności jest abstrakcją matematyczną i jej sens nie zależy od

systemu, w którym się ją zdefiniuje (jeśli jest równoważny maszynie Turinga) (Penrose 1994, s. 66). Teza ta nie została nigdy formalnie udowodniona, ale nie udało się nikomu znaleźć takiego algorytmu, dla którego nie dałoby się zaprojektować odpowiedniej maszyny Turinga (Domańska 1991).

Oczywiście tak, jak istnieją systemy równoważne (potrafiące obliczyć te same funkcje) maszynie Turinga, tak istnieją systemy, których maszyna ta jest uogólnieniem. (...) Maszynę Turinga charakteryzowały trzy cechy: skończona liczba stanów, nieskończona pamięć i nieograniczony dostęp do pamięci. Najprostsze interesujące systemy obliczeniowe to automaty stanu skończonego. Nie mają one pamięci zewnętrznej, interpretują tylko skończoną liczbę sygnałów wejściowych przechodząc ze stanu aktualnego do stanu określonego w tabeli przejścia, w której każdej akceptowanej przez automat parze «symbol wejściowy, stan» przyporządkowany jest nowy stan. System o większej mocy obliczeniowej to automat ze stosem, dysponuje on pamięcią o nieskończonej pojemności, lecz dostęp do pamięci jest ograniczony. Automat może odczytać wyłącznie symbol ostatnio zapamiętany. Aby dotrzeć do symbolu wcześniejszego trzeba odczytać wszystkie symbole zapamiętane później od niego. Następną klasą systemów obliczeniowych są maszyny von Neumanna – formalne modele komputerów. Chociaż z powodu skończonej pojemności pamięci są słabsze obliczeniowo od maszyn Turinga, w przeciwieństwie do nich są wygodne w programowaniu dzięki możliwości zapisania programu w odpowiednio wydzielonej pamięci. Rozpatrując powyższe modele badacze nauki poznawczej nie sądzą jednak, iżby maszyna Turinga czy von Neumanna były dobrymi modelami ludzkiego umysłu (Green i in. 1996).

Powstało w związku z tym na gruncie nauki o sztucznej inteligencji wiele systemów manipulujących reprezentacjami systemów obliczeniowych, których twórcy starali się udowodnić ich efektywność dla rozwiązywania problemów właściwych ludzkiej inteligencji. Wiele z tych systemów (choć nie wszystkie) wykorzystano w nauce poznawczej dla opisanego pewnych procesów poznawczych, stworzono też systemy nowe. Systemy te dzielą się na dwie podstawowe klasy: systemy symboliczne (bazujące na analogii umysłu) i systemy koneksjonistyczne (bazujące na analogii mózgu).

## 4.2. Symbolizm

Klasa systemów symbolicznych obejmuje systemy manipulujące symbolami.

W terminologii twórcy semiotyki C. S. Peirce'a pojęcie symbolu stanowi jedną z trzech podklas pojęcia znak. Znak to obiekt, który przywodzi na myśl inny obiekt, który to obiekt znak tym samym oznacza. Może tak się dziać wg Peirce'a z trzech przyczyn: znak może przypominać wskazany obiekt (wtedy Peirce zowie go 'znakiem ikonicznym'), znajduje się w znanej nam relacji do wskazanego obiektu (znak umotywowany), lub oznacza (desygnuje) ten obiekt na mocy ogólnie przyjętej konwencji (symbol) (Gregory 1996, s. 764). Znak ikoniczny to np. uproszczony rysunek przedmiotu – ikona drukarska w edytorze tekstu przypomina urządzenie, które przeprowadza drukowanie. Znak umotywowany to np. termometr – dzięki skali znamy relację między długością słupka rtęci a temperaturą otoczenia. Symbol nie wykazuje tych dwóch cech. Uzyskuje znaczenie wyłącznie na mocy interpretacji. Materialny nośnik symbolu nie wykazuje podobieństwa do obiektu oznaczanego, symbol zastępuje obiekt jedynie na mocy umowy między używającymi go osobami.

Dla procesów obliczeniowych znaczenie ma nie pojedynczy symbol, ale system symboli powiązanych wzajemnie relacjami. Według twórców Tezy o Fizycznym Systemie Symboli,

Alana Newella i Herberta Simona, procesy obliczeniowe to właśnie zachowanie się systemów symboli. Zgodnie z Tezą o Fizycznym Systemie Symboli system symboli ma własności konieczne i wystarczające do inteligentnego działania. Teza ta może być wykazana wyłącznie empirycznie poprzez badanie systemów budowanych w ramach nauki o sztucznej inteligencji jako fizyczne systemy symboli (Rich i Knight 1991, ss. 6-7). Jeśli zachowują się one inteligentnie, czyli rozwiązują problemy rozwiązywane przez człowieka, to system symboli wystarcza dla realizacji inteligencji. W nauce poznawczej ważny jest też aspekt wyjaśniania przez fizyczny system symboli istoty inteligencji człowieka. Jeśli system taki zapewnia najlepszą teorię inteligentnego zachowania człowieka, oznacza to, że prawdopodobnie procesy poznawcze człowieka mają charakter symboliczny.

Fizyczny system symboli posiada cztery składowe: pamięć, symbole, operacje na symbolach i metody interpretacji symboli. Pamięć to zdolność przechowywania przez system odpowiednich struktur symboli [*symbol structures*] zawierających symbole atomowe [*symbol tokens*], dająca możliwość dostępu do innych przechowywanych struktur. System symboli nie przetwarza naraz wszystkich pamiętanych przezeń struktur, gdyż jako system fizyczny podlega fizycznym ograniczeniom. Przetwarza tylko te struktury, które prowadzą do osiągnięcia celu. Tworzenie nowych struktur bądź odszukiwanie w pamięci struktur dotychczas niewykorzystywanych jest możliwe dzięki zdolności systemu do dokonywania operacji na symbolach, czyli ich przekształcania. Proces interpretacji struktur symbolicznych polega zaś na tym, że niektóre struktury mogą wywoływać określoną sekwencję operacji na symbolach. Takie struktury to np. procedury lub programy. Dzięki temu, że wiedza systemu oddzielona jest od sposobów jego zachowania, system może optymalizować swoje zachowanie nie zmieniając posiadanej wiedzy. Sam system symboli nie zapewnia jeszcze reprezentacji świata zewnętrznego. Dopiero system zbudowany na architekturze opartej o system symboli i wzbogacony o sensoryczne i motoryczne interfejsy ze światem zewnętrznym jest, według Newella i Simona, zdolny do sprawnego funkcjonowania w tym świecie (Newell i in. 1989, ss. 104-107).

Wzorcowym modelem ludzkiej architektury poznawczej jest tzw. model standardowy [*the standard model*]. W modelu tym występują dwa rodzaje pamięci: pamięć długotrwała o (teoretycznie) nieograniczonej pojemności i pamięć krótkotrwała o pojemności kilku elementów<sup>9</sup>. W pamięci krótkotrwałej, o krótkim czasie dostępu, przechowywane są struktury symboli aktualnie przetwarzane. W pamięci długotrwałej, o dłuższym czasie dostępu i jeszcze dłuższym czasie zapisu, znajduje się cała wiedza systemu. Elementy pamięci krótkotrwałej nie muszą być przechowywane przez osobny podsystem, ale wyróżniane poprzez nadanie cechy aktywności strukturom w pamięci długotrwałej. W takim przypadku oba rodzaje pamięci zapisane są w ten sam sposób, ale dostęp do struktur aktywnych jest szybszy niż do struktur nieaktywnych (Simon i Kaplan 1989, s. 9).

Pogląd na to, w jaki sposób wybierane są struktury do przetwarzania, dzieli badaczy z kręgu symbolicznego na dwie grupy. Pierwsza grupa, najczęściej badacze, których myśl zakorzeniona jest w filozofii i lingwistyce, uważa, że wiedza zapisana w pamięci ma postać struktur o charakterze językowym (np. zdań), a operacje odpowiedzialne za jej przetwarzanie mają postać reguł wnioskowania. Druga grupa, przeważnie psychologowie i badacze sztucznej inteligencji, rozważa wiedzę jako modele umysłowe, a ich przetwarzanie ma postać heurystycznego przeszukiwania przestrzeni tych modeli. Systemy oparte wyłącznie na wnioskowaniu stosowane są np. w symulacji myślenia dedukcyjnego w języku Prolog czy w

<sup>9</sup> Model pamięci krótkotrwałej przedstawił w 1956 r. G. Miller na konferencji w MIT. Jest to klasyczny model 7±2, gdzie cyfry wskazują na pojemność jednostek informacji. Przyp. M. Kasperski.

rozpoznawaniu poprawności gramatycznej tekstu [*parsing*]. Przeszukiwanie stosowane jest m.in. w symulacji rozwiązywania problemów i myślenia indukcyjnego (Tamże, s. 14). Podział powyższy jest jedynie modelowy. Zgodnie z badaniami empirycznymi myślenie odbywa się raczej poprzez dowolne stosowanie wielu reguł zawierających informację semantyczną. Choć bardziej przypomina heurystyczne szukanie odpowiedniego modelu niż ścisłą logiczną dedukcję, to ostry podział między poglądem, że myślenie oparte jest na logice/języku a wizją myślenia jako heurystycznego poszukiwania jest w takim (empirycznym) ujęciu rozmyty. Przeszukiwanie przestrzeni modeli nie polega bowiem na stworzeniu kompletnej przestrzeni i jej przeglądaniu. Modele są generowane z pewnej liczby modeli początkowych poprzez stosowanie operatorów zgodnie z pewnymi heurystykami. W myśleniu dedukcyjnym, aby przy przekształceniu zdań zachować ich prawdziwość, stosuje się niewielką liczbę pierwotnych reguł wnioskowania (np. w systemie dedukcji naturalnej Słupeckiego-Borkowskiego jest siedem: reguła odrywania oraz 3 pary reguł dołączania i opuszczania koniunkcji, alternatywy i równoważności (Borkowski 1991, s. 31)). Mała liczba reguł ułatwia weryfikację poprawności wnioskowania. Dodając do systemu logicznego wiele innych reguł wnioskowania w tym reguł empirycznych, duża część wiedzy na temat domeny zapisana zostaje w regułach wnioskowania, a nie w postaci deklaratywnej. W ten sposób system bardziej przypomina zbiór operatorów do generowania i przeszukiwania przestrzeni rozwiązań niż system logiki. System taki jest jednak nie mniej logiczny od pierwotnego systemu przeprowadzającego rozumowanie (Simon i Kaplan 1989, ss. 18-20).

Podstawowa klasa systemów symbolicznych to systemy regułowe [*production systems*], zaproponowane przez Newella i Simona (Newell i Simon 1972) jako architektura symboliczna procesów poznawczych. Systemy te pozwalają realizować zarówno wnioskowanie za pomocą reguł, jak i przeszukiwanie przestrzeni stanów (ciągów wnioskowań) za pomocą operatorów. Integrują zatem dwa wspomniane poglądy na przebieg myślenia. System regułowy definiuje się (Rich i Kight 1991, s. 36) jako system, na który składają się:

- zbiór reguł w postaci implikacji  $A \rightarrow B$ , gdzie A to zbiór warunków wystarczających do stosowalności reguły, a B to zbiór operacji, jakie system podejmuje, gdy dana reguła zostaje zastosowana,
- jedna lub więcej baz wiedzy, o dowolnej strukturze i zawierających dowolne informacje, przy czym niektóre z baz mogą być przypisane do systemu na stałe, a inne tylko dla potrzeb rozwiązywania bieżącego problemu,
- strategia pozwalająca wybrać jedną regułę spośród zbioru reguł, których warunki stosowalności zostały spełnione,
- system stosujący reguły.

Moc obliczeniowa systemu regułowego jest równa mocy uniwersalnej maszyny Turinga, jako że właśnie system oparty na regułach a zaproponowany przez logika Emila Posta jest jedną z równoważnych definicji obliczalności (Johnson-Laird 1988, s. 162). Systemy regułowe są zresztą implementowane często w języku LISP, który, jak wspomniano powyżej, zawiera rachunek  $\lambda$ .

Dwa podstawowe sposoby wnioskowania implementowane w systemach regułowych oparte są na logice klasycznej. Obydwa sposoby wykorzystują regułę odrywania (*modus ponens*) mówiącą, że ze zdania (jeśli A to B) i A) wynika B, gdzie A i B to zdania danego systemu logiki klasycznej. Wnioskowanie w przód [*forward reasoning*] polega na stosowaniu reguły odrywania do znanych reguł i wiedzy systemu tak długo, aż zostanie oderwany postawiony

cel (hipoteza). Wnioskowanie to lawinowo zwiększa objętość wiedzy, co przy niewielkiej jej objętości jest zaletą, lecz przy wielkości zbyt dużej grozi eksplozją kombinatoryczną. Drugi sposób wnioskowania, wnioskowanie wstecz [*backward reasoning, goal-directed reasoning*], polega na próbie wykazania prawdziwości przesłanek reguły, której konkluzja stanowi (cel) hipotezę wnioskowania. Jeśli przesłanki występują w bazie wiedzy to cel zostaje osiągnięty. Jeśli któraś z przesłanek nie jest znana systemowi należy założyć ją jako hipotezę tymczasową i próbować wykazać stosując ponownie metodę wnioskowania wstecz. Metoda ta nie powiększa lawinowo bazy wiedzy, ale stosując ją można jednocześnie wnioskować tylko jedną hipotezę. Możliwe jest także wnioskowanie mieszane, w którym na podstawie pewnych metareguł dokonuje się przełączenia między wnioskowaniem w przód i wstecz, zależnie od przebiegu wnioskowania (Mulańska 1996, ss. 81-86).

Powyższe metody wnioskowania możemy także rozumieć jako metody konstruowania grafu przestrzeni rozwiązań. Wnioskowanie w przód odpowiada tworzeniu grafu z początkowych założeń jako węzła początkowego i generowaniu następnych węzłów w konkluzji reguł, których przesłankę spełnia węzeł początkowy. Następnie proces jest iterowany: do nowo wygenerowanych węzłów stosowane są reguły o przesłankach spełnianych przez te węzły. Jeśli wygenerowany zostanie cel (hipoteza) to droga w grafie prowadząca od węzła początkowego do celu jest znalezionym rozwiązaniem problemu. Wnioskowanie wstecz odpowiada generowaniu grafu z hipotezy jako węzła początkowego i stosowaniu tych reguł, których konkluzją jest hipoteza (czyli tych, które doprowadziłyby nas do rozwiązania, gdyby ich przesłanki były prawdziwe). Przesłanki tychże reguł tworzą nowe węzły. Proces jest iterowany do nowych węzłów. Jeżeli uzyskany zostanie węzeł stanowiący początkowe założenia droga w grafie prowadząca do węzła początkowego do celu analogicznie jak w poprzedniej metodzie jest znalezionym rozwiązaniem problemu (Rich i Knight 1991, ss. 178-179).

System regułowy ma modelować inteligencję. Musi więc minimalizować czas rozwiązania problemu (aby szybko zaadaptować się do warunków środowiska) oraz liczbę możliwych dróg prowadzących do rozwiązania problemu (gdyż pamięć systemu jest ograniczona). Strategie przyspieszające działanie systemu i minimalizujące ilość danych, jaką w danym momencie system musi przechowywać w pamięci, polegają na odpowiednim wyborze pewnych węzłów grafu spośród wszystkich aktywnych węzłów (strategie przeszukiwania grafów), a następnie na wyborze odpowiednich reguł spośród wszystkich możliwych do zastosowania dla wybranego węzła (metody sterowania wnioskowaniem).

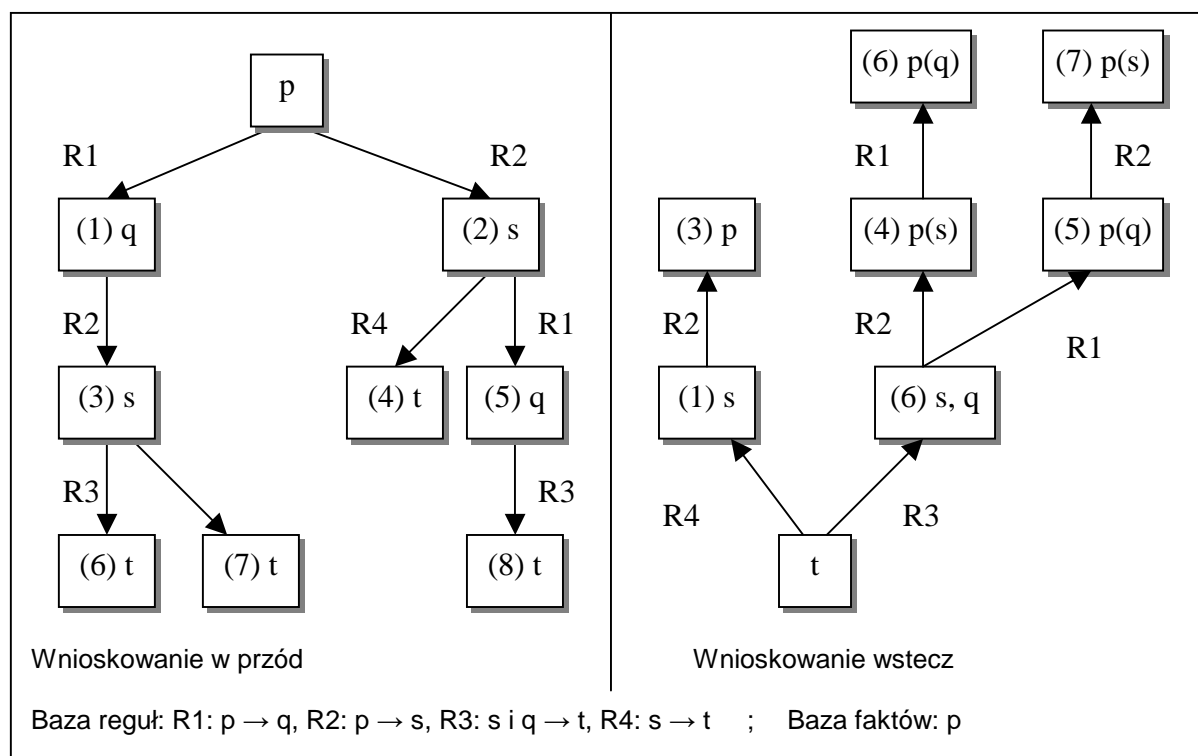
Wśród strategii przeszukiwania grafów wyróżniamy strategie ślepe, nie wykorzystujące wiedzy o rozwiązywanym problemie, z których najbardziej znane to:

- strategia w głąb [*depth-first*] polegająca na generowaniu kolejno gałęzi grafu; gdy aktualnie generowana gałąź prowadzi do węzła, dla którego nie istnieje odpowiadająca reguła zostaje generowana kolejna gałąź aż do osiągnięcia węzła stanowiącego cel,
- strategia wszerz [*breadth-first*] polegająca na generowaniu wszystkich gałęzi równolegle, z tym, że wybierane są węzły leżące najbliżej węzła początkowego,
- strategia zachłanna [*hill-climbing*] polegająca na wybraniu jednego optymalnego węzła (według zadanych kryteriów) i badaniu następnie tylko jego następników – powrót do poprzednich węzłów jest niemożliwy.

Najpopularniejszą strategią heurystyczną, czyli wykorzystującą pewną wiedzę o problemie, jest strategia 'najpierw najlepszy' [*best-first*], polegająca na określeniu dla wszystkich



węzłów pewnej funkcji heurystycznej zależnej od zbieżności węzła do celu, najmniejszego kosztu drogi od węzła początkowego i najmniejszej złożoności obliczeniowej procesu przeszukiwania i wybieraniu węzła o optymalnej wartości funkcji heurystycznej (Mulawka 1996, ss. 75-78).



Rysunek 1. Schemat wnioskowania w przód i wstecz (oprac. własne).

Powyższy schemat dostarcza przykładu, w jaki sposób stosowanie reguł możemy traktować jako przeszukiwanie grafu rozwiązań. Mamy dane: 4 reguły oraz wiedzę na temat jednego faktu „p”. Cel stanowi uzasadnienie faktu „t”. Graf pokazuje, jakie reguły były możliwe do zastosowania w każdej fazie wnioskowania i wiedzę, o jakim fakcie dzięki nim uzyskano. Stosując wnioskowanie w przód możemy użyć dwóch reguł: R1 i R2, gdyż „p” spełnia ich przesłankę. Jednakże zastosowanie R2 wprowadza do bazy faktów „s”, do którego możemy zastosować R4 i uzyskać rozwiązanie „t”. Inne drogi rozumowania także prowadzą do oczekiwanej konkluzji, wymagają jednak dłuższego wnioskowania. Przeszukując graf możliwych wnioskowań korzystając ze strategii wszerz uzyskalibyśmy „t” w czterech krokach (1), (2), (3), (4). Zastosowanie strategii w głąb doprowadziłoby nas do „t” w trzech krokach (1), (3), (6) lub (7). Dla strategii zachłannej założymy, że funkcja wyboru węzłów przyjmuje tym większą wartość, im uzyskana w węźle literka jest w alfabecie bliżej „t”. Zatem uzyskano by rozwiązanie w dwóch krokach: (2) i (4). Przy innej postaci funkcji strategia ta mogłaby się powieść w trzech krokach. Przykładowy problem jest jednak bardzo prosty. Dla innego problemu, przy innej bazie wiedzy i źle wybranej funkcji wyboru węzłów strategia ta mogłaby w skrajnym przypadku w ogóle nie doprowadzić do rozwiązania.

Prześledźmy teraz przebieg wnioskowania wstecz. System próbuje wykazać „t” sprawdzając czy przesłanki dwóch reguł zawierających w konkluzji „t” (i tylko „t”): R3 i R4 znajdują się w bazie wiedzy. Stwierdzając, że warunek taki nie zachodzi, system próbuje wykazać przesłanki R3 i R4. Dla R3 przesłankę stanowi „s”, fakt ten jest jednocześnie konkluzją R2,

której przesłanka „p” znajduje się w bazie wiedzy. A zatem „t” zostaje wykazana. Jeśli zaczniemy od analizy przesłanek R4 uzyskanie oczekiwanego wyniku zabierze więcej czasu. R4 ma bowiem dwa fakty „s” i „q” jako przesłanki, ponieważ łączy je koniunkcja musimy dla obu znaleźć reguły, mające jako przesłankę „p” (i tylko „p”): R1 i R2. W tym przypadku strategia w głąb prowadzi do wyniku po dwóch krokach (1) i (3), a strategia wszerz po trzech (1), (2) i (3). System mógłby zastosować także wnioskowanie mieszane. Stosując wnioskowanie w przód i regułę R2 uzyskałby „s”. Mógłby więc wnioskować wstecz za pomocą reguły R4 – jej przesłanka jest już w bazie wiedzy.

Metody wyboru reguły spośród reguł możliwych do zastosowania polegają m.in. na wyborze reguł ostatnio uaktywnionych (strategia świeżości), blokowaniu reguł ostatnio wykorzystywanych (strategia blokowania) oraz na wyborze reguł o największej liczbie przesłanek (strategia specyficzności) (Mulawka 1996, s. 83).

Klasa systemów regułowych obejmuje wiele systemów rozwiązujących konkretne problemy, ale zawiera też rodzinę systemów ogólnych:

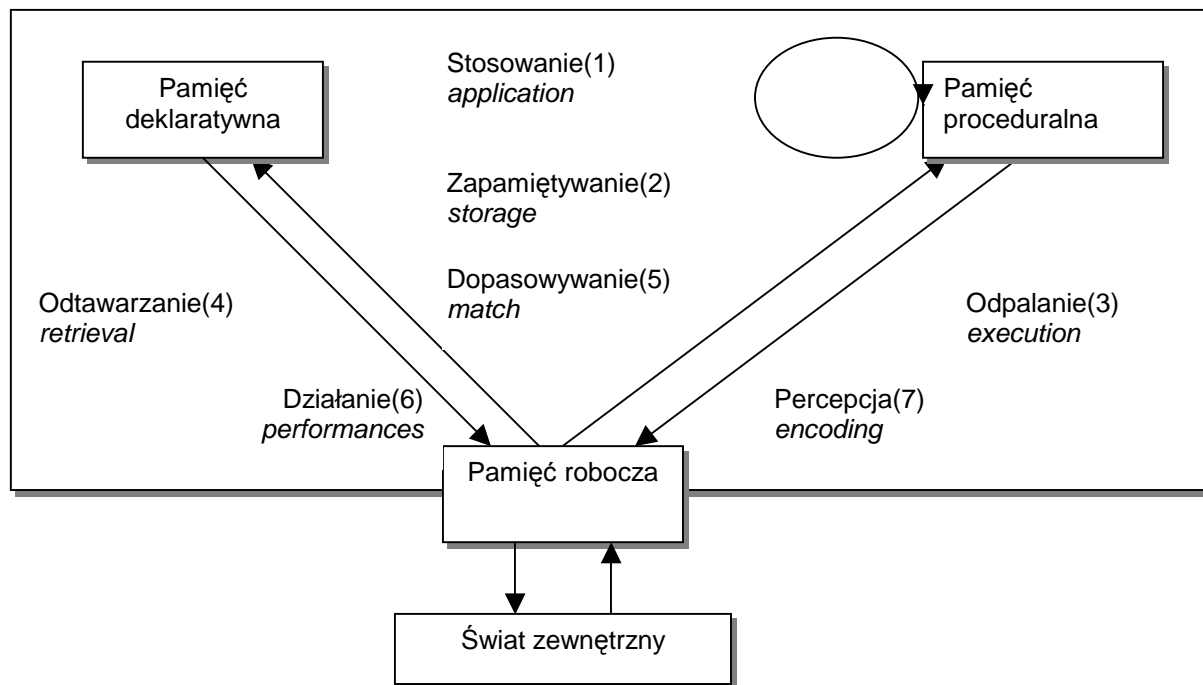
- języki do tworzenia systemów regułowych, jak np. ACT\*<sup>10</sup>,
- systemy szkieletowe zapewniające środowisko do budowy systemów ekspertowych,
- ogólne architektury procesu rozwiązywania problemów zbudowane na bazie zbioru hipotez poznawczych odnoszących się do tego procesu, jak np. system SOAR (Rich i Knight 1991, s. 36).

Chociaż wiele systemów regułowych to systemy ekspertowe, jednak pamiętać należy, że pojęcie systemu regułowego nie pokrywa się z pojęciem systemu ekspertowego. Istotą systemu ekspertowego jest jego funkcja: rozwiązywanie specjalistycznych problemów wymagających profesjonalnej ekspertyzy (Mulawka 1996, s. 20). Istotą systemu regułowego jest jego struktura – zbiór reguł i operacje manipulowania nimi. Chociaż ok. 85% systemów ekspertowych to systemy regułowe (tamże, s. 31), istnieją także systemy ekspertowe oparte np. na sieciach neuronowych. I na odwrót, obok istniejących w nauce o sztucznej inteligencji systemów regułowych realizujących funkcję eksperta występują na gruncie nauki poznawczej wspomniane powyżej systemy regułowe stanowiące ogólny model wnioskowania. Dla nauki poznawczej szczególne znaczenie mają oczywiście te architektury symboliczne, które – przynajmniej w zamierzeniach swoich autorów – opisują i wyjaśniają procesy poznawcze człowieka. Wśród nich warto omówić dwie wcześniej wspomniane i szeroko w nauce poznawczej opisywane: ACT\* i SOAR.

System ACT\* stworzony przez J. Andersona to pierwszy model procesów poznawczych o wystarczająco pełnej i szczegółowej architekturze. Składa się z pamięci krótkotrwałej oraz z dwóch rodzajów pamięci długotrwałej: pamięci deklaratywnej, w postaci sieci semantycznej, i pamięci proceduralnej, jako zbioru reguł. Zarówno węzłom sieci jak i regułom przypisany jest atrybut aktywności ewoluowany w trakcie przetwarzania. ACT\* posiada kilka podstawowych typów węzłów sieci semantycznej, m.in. wartości liczbowe, łańcuchy znakowe, tablice. Reguły powodują uaktywnienie się symboli w swoich konkluzjach. Pamięć krótkotrwałą stanowią: aktywny fragment pamięci deklaratywnej, elementy deklaratywne uzyskane w wyniku odpalania aktywnych reguł oraz elementy uzyskane dzięki percepcji. Pamięć ta nie jest więc osobnym zbiorem, ale podzbiorem zbioru danych deklaratywnych. Aktywacja rozchodzi się w sieci automatycznie jako funkcja wagi przypisanej elementom

<sup>10</sup> Więcej na temat ACT patrz m.in. w: P. Czarnecki, *Koncepcja umysłu w filozofii D. Dennetta*, rozdz. Świadomość, <http://www.kognitywistyka.net>. Przyp. M. Kasperski.

pamięci i celu wnioskowania. Waga elementów wzrasta wraz z użyciem danego elementu przez system. W procesie wyboru reguł do dopalenia, czyli uzyskania aktywności, wygrywają reguły i węzły sieci o największej wartości wag. Aktywacja elementów nieużywanych wygasa (Newell i in. 1989, s. 110).

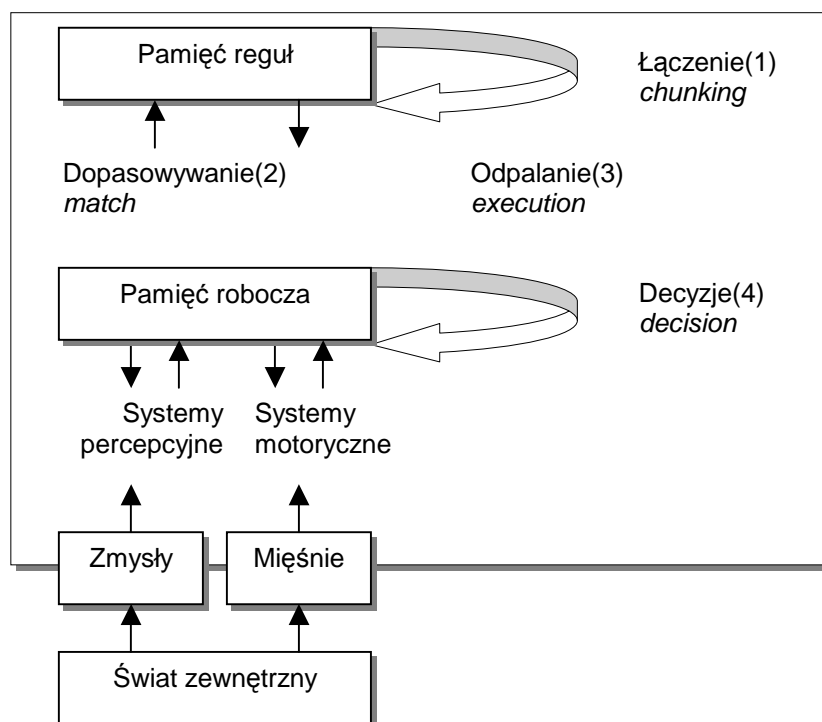


Rysunek 2. Schemat architektury poznawczej ACT\*. (Newell i in. 1989, s. 110).

Dopasowując (5) reguły z pamięci proceduralnej do aktywnych symboli system uzyskuje zbiór reguł możliwych do odpalenia (3). Odpalając regułę o największej wadze aktywuje symbole będące jej konkluzją. Następnie system uzyskuje (4) elementy sieci semantycznej do pamięci krótkotrwałej poprzez rozchodzenie się aktywacji w sieci od uaktywnionych przez odpalenie reguł symboli. W wyniku przetwarzania [encoding] za pomocą reguł działających niezależnie od reszty systemu danych pochodzących z percepcji (7) do pamięci roboczej dodawane są konkluzje tychże reguł. Istnieje stałe prawdopodobieństwo, że dane tak uzyskane zostaną zapamiętane (2) w pamięci deklaratywnej. System ma zaimplementowane także procedury uczenia się nowych reguł na podstawie poprzednich zachowań. W trakcie rozwiązywania problemu może zostać utworzona reguła, która prowadzi od założeń do celu. Zawiera ona w sposób niejawni cały przebieg wnioskowania. Poprzez jej zastosowanie (1) problem zostaje rozwiązany i system oddziałuje (8) na świat zewnętrzny (Newell i in. 1989, ss. 112-115).

SOAR to system, który oprócz procesów centralnych (które w ACT\* dominują) ma także rozwinięte moduły odpowiadające za percepcję i motorykę. Ma jeden rodzaj pamięci długotrwałej – system regułowy, używany do przechowywania zarówno wiedzy proceduralnej jak i deklaratywnej. System ten ma postać pamięci skojarzeniowej: uruchomienie reguły prowadzi do symboli występujących po jej prawej stronie. Występuje jeden rodzaj elementów pamięci: zbiór atrybutów i ich wartości. Pamięć robocza jest oddzielona i zawiera: hierarchię celów, informacje z nią związane, preferencje o tym, co powinno być wykonane oraz informacje percepcyjne i motoryczne. Krótkotrwałość pamięci

robotycznej zapewniona jest powiązaniem jej zawartości z aktualnym celem. Jeśli zmienia się cel, zmienia się też jej zawartość. Interakcja ze światem zewnętrznym odbywa się poprzez jeden lub więcej wyodrębnionych modułów percepcyjnych i jeden lub kilka modułów motorycznych. Wszystkie problemy formułowane są w przestrzeni stanów a reguły generujące stany odpalane są równoległe: decyzje dotyczą przestrzeni stanów, stanów i operatorów (Newell i in. 1989, ss. 110-112).



Rysunek 3. Schemat architektury poznawczej SOAR (Newell i in. 1989, s. 111).

System przetwarza informacje rozpoczynając od zawartości pamięci roboczej, w której m.in. znajduje się aktualny cel i informacje z nim związane. Wybór operatorów (czyli ciągów reguł) do rozwiązywania problemu odbywa się w dwuetapowym cyklu decyzyjnym. Najpierw system odczytuje zawartość pamięci długotrwałej równoległe (2 i 3). Podczas tego procesu jedne odczyty mogą powodować następne. Dzieje się tak aż do ustabilizowania procesu. Pomiędzy odczytanymi elementami pamięci znajdują się także preferencje dotyczące wyboru operatorów. W drugiej fazie zostaje podjęta decyzja co do wyboru właściwych spośród odczytanych operatorów(4) i system przechodzi do następnego cyklu. Jeżeli SOAR nie jest w stanie podjąć decyzji, problem wyboru operatorów traktuje jako tymczasowy cel i przystępuje do rozwiązania tak zadanego problemu w sposób opisany powyżej. Proces uczenia się polega na łączeniu reguł podobnej zasadzie jak w systemie ACT\* (Newell i in. 1989, ss. 114-117).

Różnice pomiędzy oboma systemami są mniejsze niż wynika to z opisu, np. chociaż SOAR ma jeden moduł pamięci długotrwałej, a ACT\* dwa, to moduł systemu SOAR posiada cechy obydwu modułów ACT\*, reguły wpisane są bowiem w architekturę sieci semantycznej (tamże, s. 120).

Obok opisanego powyżej podejścia polegającego na traktowaniu procesów myślenia jako dowolnego stosowania wielu reguł zawierających informację semantyczną, czyli procesu

zbliżonego do przeszukiwania przestrzeni modeli umysłowych na bazie reguł (zdań), zaproponowano także podejście oparte na przetwarzaniu obrazów umysłowych [*mental images*], czyli podejście odwołujące się do geometrii i analizy przestrzennej. Człowiek, w tym ujęciu, tworzy dwuwymiarowy obraz analizowanej sytuacji na swoim wewnętrznym „ekranie” i przegląda ten obraz oraz przetwarza go stosując takie operacje jak np. powiększenie czy obrócenie. Są to więc zupełnie inne operacje niż w systemach regułowych. Eksperymenty (przeprowadzone przez S. Kosslyna) pokazały, że w przypadku modelowania wyobrażeń taka reprezentacja i jej przetwarzanie mają silne potwierdzenie empiryczne. M.in. wyobrażając sobie mapę jakiegoś obszaru człowiek orzeka o relacji pomiędzy dwoma jej elementami w czasie wprost proporcjonalnym do odległości między nimi na owej mapie (Green i in. 1996, s. 14).

### 4.3. Koneksjonizm

Dużą popularność od lat osiemdziesiątych zyskało w nauce poznawczej odmienne od symbolicznego podejście do natury procesów obliczeniowych – podejście koneksjonistyczne. Wykorzystywane początkowo głównie do symulacji procesów percepcji obecnie ma zastosowanie we wszystkich domenach nauki poznawczej. W odróżnieniu od systemów symbolicznych inspirowanych funkcjonowaniem komputera i/lub ideą umysłu, modele koneksjonistyczne wzorowane są na budowie i funkcjonowaniu mózgu.

Mózg to struktura składająca się z ogromnej liczby (ok. 100 miliardów) połączonych w sieć (ok. 100 bilionów połączeń) prostych i podobnych do siebie elementów zwanych neuronami. Funkcjonowanie owej sieci polega na dynamicznym przetwarzaniu za pomocą nieskomplikowanych operacji (dodawanie, odejmowanie i mnożenie algebraiczne oraz suma i iloczyn logiczny) sygnałów elektrochemicznych otrzymywanych przez pojedynczy neuron od innych neuronów w jeden sygnał przekazywany dalej do następnych neuronów. Mózg przetwarza informacje równoległe, tzn. w określonym czasie pobudzany jest nie jeden neuron, ale określone grupy neuronów (Wróbel 1997a). Pobudzenie to jest prawdopodobnie zsynchronizowane i przebiega w fazach odpowiadających częstotliwościom fal elektroencefalograficznych, np. częstotliwość oscylacyjnej aktywności grupy neuronów kory wzrokowej kota zawiera się w paśmie  $\gamma$  (30-90Hz) (Wróbel 199b, s. 472).

Analogicznie do pracy mózgu, cechą systemów koneksjonistycznych jest równoległe przetwarzanie informacji przez wiele połączonych w sieć prostych przetworników (analogonów neuronów). Procesy obliczeniowe w sieci są prowadzone jako proste interakcje pomiędzy połączonymi ze sobą przetwornikami. Interakcje te polegają na przesyłaniu określonych wartości liczbowych wzdłuż połączeń między nimi. Obliczenia prowadzone są równoległe, co oznacza, że informacja jest przetwarzana przez dużą liczbę przetworników w tym samym czasie. Wiedza systemu jest zapisana niejawnie w strukturze połączeń między przetwornikami. To, które jednostki przetwarzające są ze sobą połączone oraz siła tego połączenia, decyduje o tym, jaką reprezentację świata zewnętrznego posiada system koneksjonistyczny. Równoległość obliczeń, struktura sieci prostych przetworników oraz zapisanie wiedzy systemu w tej strukturze definiują systemy koneksjonistyczne (Rumelhart 1989, s. 134-136). Trzy powyższe warunki muszą występować równocześnie, gdyż np. także symboliczny system SOAR wykazuje cechę równoległości obliczeń (choć brak mu dwu cech pozostałych).

Należy pamiętać, że koneksjonizm korzysta z wiedzy neurofizjologicznej tylko na zasadzie analogii, systemy koneksjonistyczne nie symulują pracy sieci neuronowych i nie biorą pod

uwagę biologicznych ograniczeń swoich naturalnych odpowiedników. Koneksjonizm jest motywowany przez zjawiska poznawcze (a nie neurofizjologiczne) i rządzony przez ograniczenia obliczeniowe. Systemy koneksjonistyczne zwane są demonstracjami [*demonstrations*] systemu nerwowego, jako że nie dokonują predykcji jego funkcjonowania, a tylko wskazują pewien możliwy – zresztą bardzo uproszczony – sposób jego działania. W odróżnieniu od nich, modele rzeczywistych sieci neuronowych starają się dokładnie opisywać dynamiczną automodyfikację neuronu i nie upraszczają tego procesu jako obliczania wyniku w sztucznym neuronie (Sejnowski i Churchland 1989, s. 342).

System koneksjonistyczny składa się z siedmiu głównych elementów (Rumelhart 1989, s. 136-137):

- zbioru jednostek przetwarzających,
- stanu aktywacji określonego na tym zbiorze,
- funkcji aktywacji określonej dla każdej jednostki, która stanowi jej aktywacji przyporządkowuje wartość sygnału wyjściowego,
- struktury połączeń między jednostkami,
- reguły, która przyporządkowuje wartościom sygnałów wejściowych i stanowi aktywacji jednostki nową wartość stanu aktywacji,
- reguły uczenia się sieci,
- środowiska, w którym sieć operuje.

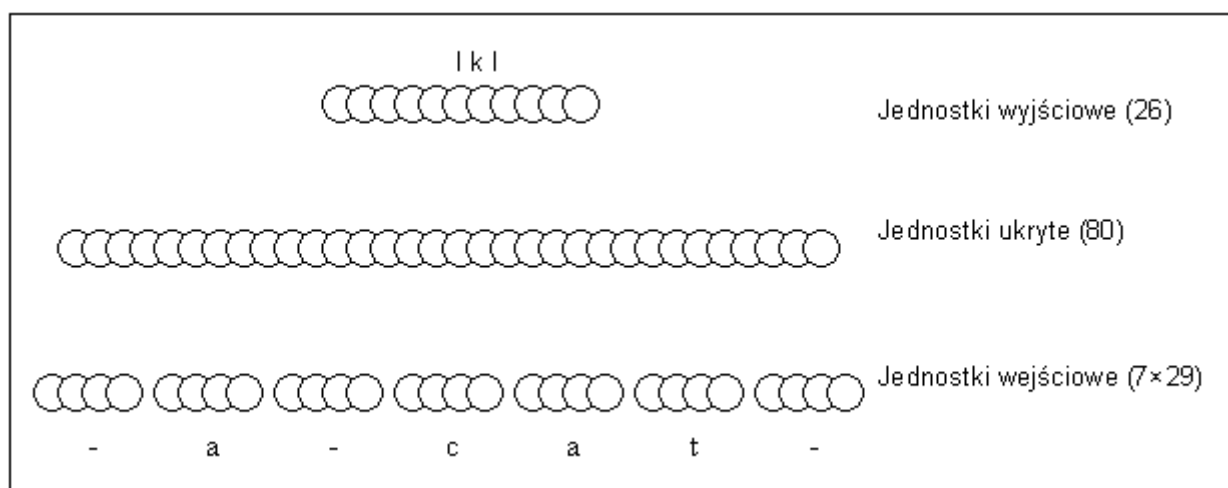
W zależności od tego, w jaki sposób systemy koneksjonistyczne reprezentują obiekty ze środowiska, możemy je podzielić na systemy przetwarzania rozproszonego [*distributed processing*] lub systemy lokalistyczne [*localist*]. W systemach przetwarzania rozproszonego pojedyncza jednostka jest tylko abstrakcyjnym elementem nie reprezentującym czegokolwiek, reprezentacja jest określona dopiero na zbiorze tych jednostek. W odróżnieniu, element systemu lokalistycznego reprezentuje pewne pojęcie lub inny posiadający znaczenie obiekt (tamże, s. 138).

Wyróżnia się w nauce poznawczej trzy typy systemów koneksjonistycznych: sieci jednokierunkowe, sieci rekurencyjne i sieci asocjacyjne (jako szczególnie ważną ze względu na swoją funkcję podklasę sieci rekurencyjnych) (Green i in. 1996, s. 35). Kryterium podziału systemów koneksjonistycznych na jednokierunkowe i rekurencyjne stanowi kierunek przepływu informacji w sieci.

W sieciach jednokierunkowych informacja przepływa wyłącznie od elementów wejściowych do elementów wyjściowych. Węzły sieci zazwyczaj zgrupowane są w warstwy, przy czym ze względu na swoje ograniczenia obliczeniowe sieci jednowarstwowe (o warstwie wejściowej i jednej warstwie przetwarzającej) nie są wykorzystywane w nauce poznawczej. W sieciach wielowarstwowych występuje warstwa wejściowa, co najmniej jedna warstwa ukryta i warstwa wyjściowa. Połączenia występują tylko pomiędzy elementami warstw sąsiadujących. Proces obliczeniowy w sieci jednokierunkowej przebiega w dwóch fazach. Najpierw polega na nauczeniu się za pomocą reguły uczenia odpowiedniej reakcji na sygnały wejściowe. Uczenie się sieci to ustalenie odpowiedniej wartości wag połączeń, czyli określenie struktury połączeń. Zostaje ona ustalona tak, aby na dany bodziec sieć reagowała pojawieniem się na jej wyjściużądanego sygnału. Druga faza polega na dostarczeniu nauczonej już sieci danych wejściowych i poprzez przetwarzanie tych danych w kolejnych warstwach przez elementy, zgodnie z algorytmem przedstawionym wcześniej, uzyskiwaniu sygnału na wyjściu. Istotną cechą takiego procesu obliczeniowego jest umiejętność przyporządkowywania

znieskształconym, ale podobnym do danych modelowych, danym wejściowym właściwych sygnałów wyjściowych. Inną ważną cechą sieci jest umiejętność uogólniania, czyli właściwa reakcja sieci na dane, z którymi nigdy wcześniej się nie spotkała, dzięki wiedzy sieci zapisanej w połączeniach między jej elementami.

Przykładem sieci jednokierunkowej realizującej konkretny proces poznawczy jest sieć NETtalk. Jej zadaniem jest odtwarzanie angielskiej mowy z danych wejściowych w postaci tekstu w języku angielskim. Nie jest to proste zadanie ze względu na nieregularne zasady wymowy poszczególnych zbiorów liter w tym języku. Sieć przyswoiła te zasady *implicite* jako strukturę połączeń pomiędzy jej elementami w wyniku procesu uczenia. Sieć składa się z trzech warstw. Ponieważ sposób wymowy litery zależy od liter występujących przed i za nią, warstwa wejściowa reprezentowała ciągi siedmiu kolejnych liter tekstu (trzy znaki poprzedzające, litera rozpoznawana, trzy znaki za nią) za pomocą reprezentacji lokalistycznej. Każda litera reprezentowana była jako zbiór 29 elementów (liczba znaków, które system rozpoznawał), a każdy element odpowiadał określonemu znakowi. Reprezentacją danej litery był więc wektor jednostkowy o jedynce na pozycji odpowiadającej reprezentowanej literze. Razem więc warstwa wejściowa posiadała  $29 \times 7 = 203$  elementy. Sygnał trafiał następnie do warstwy ukrytej o 80 elementach i przetwarzany był w reprezentację lokalistyczną w warstwie wyjściowej. Na reprezentację tę składało się 26 cech artykulacyjnych i sposobów akcentowania, na podstawie której cyfrowy syntezytor generował odpowiedni fonem. Ponieważ celem twórców systemu było nie tylko osiągnięcie wysokiej efektywności systemu, ale przede wszystkim modelowanie procesu generowania fonemów, analizowali oni podobieństwa i różnice systemu w stosunku do wiedzy na temat realizacji tego procesu u człowieka. Wskazywali m.in., że system pomijał, występującą u człowieka, reprezentację całych słów. Utrudniało mu to poprawną wymowę słów o nieregularnej wymowie (Crick 1997, ss. 252-258).



Rysunek 4. Schemat architektury NETtalk na przykładzie przetwarzania litery „c” w słowie „-a-cat-” w głoskę „k”. (Crick 1997, s. 254).

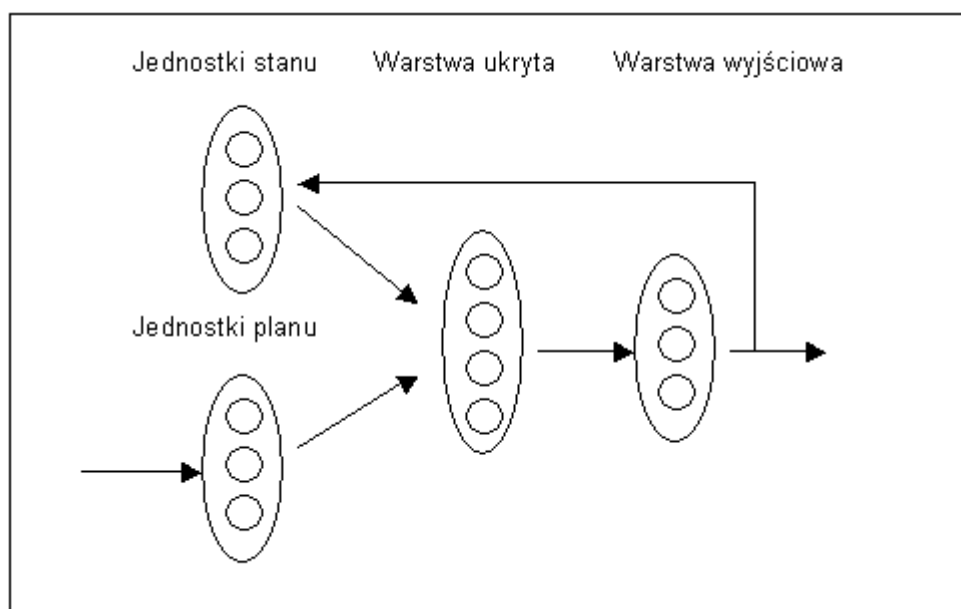
Sieci rekurencyjne to sieci posiadające sprzężenia zwrotne, tzn. taką architekturę, że sygnał wychodzący z danej jednostki jest pośrednio (przetworzony przez inne jednostki) lub bezpośrednio sygnałem wejściowym tej jednostki. Dzięki sprzężeniu zwrotnemu sygnał

wyjściowy w sieci rekurencyjnej jest funkcją zarówno aktualnego sygnału wejściowego i sygnałów poprzednich, czego nie da się uzyskać w sieciach jednokierunkowych.

Sieci asocjacyjne, zwane sieciami Hopfielda, to klasa takich sieci rekurencyjnych, które umożliwiają rekonstrukcję i rozpoznawanie wcześniej zapamiętanych wzorców na zasadzie skojarzeń na podstawie dostępnego fragmentu wzorca lub wzorca do niego podobnego. Wykorzystuje się je do modelowania pamięci skojarzeniowej. W sieciach tych nie ma wyróżnionych warstw, każda jednostka przetwarzająca połączona jest ze wszystkimi innymi jednostkami oprócz siebie samej. Połączenie między dwoma jednostkami jest symetryczne, tzn. ma taką samą siłę w obie strony. Procesy obliczeniowe polegają na nauczaniu sieci wzorcowych danych a następnie na prezentacji dowolnych danych na wejściu. Sygnał rozprzestrzenia się po sieci aż do samoczynnego uzyskania stanu stabilnego, kiedy nie zachodzą już zmiany aktywacji żadnych jednostek. Stan aktywacji określony na zbiorze jednostek przetwarzających jest przekazywany na wyjście systemu i stanowi wzorec najbardziej zbliżony do danych wejściowych (Osowski 1994, ss. 191-193).

Inne typy sieci rekurencyjnych mają architekturę, w której wydzielono warstwy, a sprzężenia zwrotne realizowane są poprzez dwukierunkowy przepływ informacji między warstwami. Dwie najpopularniejsze w nauce poznawczej architektury sieci rekurencyjnych to sieci typu Jordana i typu Elmana (Green i in. 1996, s. 36).

Sieć Jordana jest koneksjonistyczną architekturą dla procesów polegających na realizacji sekwencji podprocesów. Jej pierwsza implementacja polegała na modelowaniu sekwencji motorycznych.



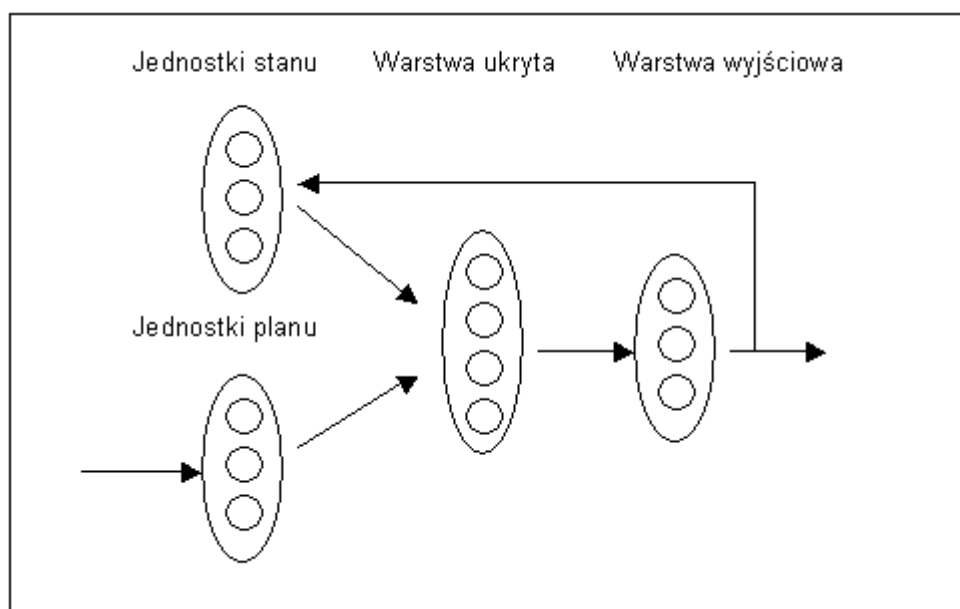
Rysunek 5. Architektura sieci typu Jordana (Green i in. 1996, s. 37).

W sieci Jordana na warstwę wejściową składają się jednostki planu, które dostarczają sygnał kodujący sekwencję procesów (np. ruchów potrzebnych do serwisu tenisowego) oraz jednostki stanu, do których dochodzi sygnał wyjściowy, odzwierciedlające dzięki temu



aktualnie wykonany podproces (np. aktualną fazę serwisu). Dzięki warstwie ukrytej sygnał zostaje przetworzony tak, aby reprezentował kolejny podproces, który zgodnie z planem należy wykonać (Jordan i Rosenbaum 1989, s. 739).

W sieci Elmana sprzężenie zwrotne dotyczy nie warstwy wyjściowej ale warstwy ukrytej. Sieć taka służyła Elmanowi do modelowania rozpoznawania zdań. W sieci Elmana z warstwy wejściowej obok sygnału wejściowego (zdania do rozpoznania) do warstwy ukrytej dostarczany jest sygnał z tzw. jednostek kontekstowych. Jednostki te stanowią pamięć przeszłych zdarzeń w systemie m.in. przetwarzanych poprzednio zdań. Jest to niezbędne dla określenia sensu zdania – inaczej zostanie zrozumiana wypowiedź „zabiję cię”, jeśli poprzednie zdania były żartem, a inaczej, gdy stanowiły groźbę. Po przetworzeniu sygnałów w warstwie ukrytej modyfikowane są stany aktywacji jednostek kontekstowych (dodana zostaje wiedza o aktualnie analizowanym zdaniu) i zarazem sygnał z tej warstwy trafia na wyjście (Rumhart 1989, ss. 154-155).



Rysunek 6. Architektura sieci Elmana (Green i in. 1996, s. 37).

#### 4.4. Symbolizm czy koneksjonizm?

Istnienie dwóch podejść: symbolicznego i koneksjonistycznego, wywołuje natychmiast pytanie o relacje między nimi. Czy nie jest tak, że wszystkie modele na gruncie jednego z podejść można z powodzeniem zastąpić ekwiwalentnymi modelami na gruncie podejścia drugiego? Wielu badaczy każdej orientacji odpowiada: „oczywiście, że tak” i przytacza natychmiast wiele zarzutów wobec strony przeciwnej.

Systemy koneksjonistyczne lepiej się uczą, wystarczy prezentować im odpowiednie wzorce, a same posiadają wiedzę, zapisaną *implicite*, potrzebną do realizacji zadania. Systemy symboliczne muszą ową wiedzę mieć zapisaną *explicite*. Ale uczenie się sieci neuropodobnych wymaga wielu prób (Rich i Kight 1992, s. 523). Kłóci się to z danymi psychologicznymi i neurofizjologicznymi, gdyż często człowiek potrafi osiągnąć wiedzę na

przykładzie jednego egzemplarza (np. przy zapamiętywaniu twarzy)<sup>11</sup>, choć istnieją procesy, które rzeczywiście wymagają długiej nauki (np. procesy motoryczne: jazda samochodem czy gra w tenisa).

Za pomocą systemów koneksjonistycznych trudniej wyjaśnić przebieg procesów poznawczych. Choć po nauczaniu sieć działa sprawnie to bardzo trudno wyjaśnić, co dzieje się w warstwach ukrytych w terminach obliczeń i reprezentacji. Z drugiej strony, człowiek w przypadku wielu procesów poznawczych też nie jest świadom reguł, które stosuje (np. widzenie, rozpoznawanie mowy) (tamże, s. 525)<sup>12</sup>.

Z kolei w systemach symbolicznych dzięki systematycznej konstrukcji reprezentacji jako struktur złożonych z prostych elementów łatwo jest uzyskać produktywność myślenia, niezbędną w realizacji wielu procesów centralnych. Na gruncie symbolizmu nie wiadomo jednak jak ta systematyczność implementowana jest w mózgu, a systemy koneksjonistyczne potrafią implementować przejrzyste reguły w postaci węzłów i połączeń między nimi. Badacze nurtu symbolicznego odpowiadają, że takie wyjaśnienie jest konieczne tylko w przypadku procesów wejścia-wyjścia (percepcji i motoryki), a nie w przypadku procesów centralnych (Green i in. 1996, s. 45).

Zarzucają z kolei koneksjonistom, że ich systemy nie są zdolne do implementacji procedur przeszukiwania rozległych przestrzeni stanów, co jest niezbędne w modelowaniu np. gry w szachy (Rich i Koght 1992, s. 525).

Wielką zaletą koneksjonizmu jest to, że korzysta w pewnym stopniu z osiągnięć neurofizjologii (stanowiąc pewien pomost w integracji *cognitive science* i *neuroscience*) i wyjaśnia niektóre (choć jak wspomniano powyżej niektórych nie wyjaśnia) zjawiska zachodzące w mózgu. Do zjawisk tych należą m.in. jego odporność na uszkodzenia<sup>13</sup> oraz równoległe i rozproszone przetwarzanie informacji<sup>14</sup> (Rumelhart 1989, ss. 134-136).

Koneksjoniści zarzucają symbolistom, że ich skłonność do wyjaśniania procesów poznawczych na poziomie syntaktycznym a nie na poziomie zbliżonym do poziomu implementacji wynika z faktu, że symbolizm opiera się na analogii komputera. Wszystkie komputery mają dosyć podobne architektury i dany proces symboliczny istotnie możemy wykonać na IBM-ie jak na Apple'u, zarówno na maszynie zbudowanej z lamp, jak z układów scalonych. Wśród systemów koneksjonistycznych istnieje jednak duża różnorodność architektur i przenoszenie funkcji jednego systemu koneksjonistycznego na drugi nie jest już takie łatwe (tamże, s. 134).

Oprócz zwolenników wyłącznie jednego podejścia istnieją w nauce poznawczej także dwie grupy badaczy akceptujących zarówno symbolizm, jak i koneksjonizm. Pierwsza grupa, opierając się na wprowadzonej przez Pylyshyna słabej równoważności procesów, twierdzi, że choć obydwa typy systemów potrafią realizować wszelkie procesy poznawcze, to niektóre z

<sup>11</sup> To nie do końca jest prawdą. Z zapamiętywaniem twarzy jest związany cały proces uczenia się, łącznie z poprzedzającym momentem wyodrębnienia przez ludzki system poznawczy czegoś takiego jak „twarz”. Więcej na ten temat patrz: J. Ingram, *Płocący dom*, rozdz. IV Twarze, ss. 169-206, Prószyński i S-ka, Warszawa 1996. Przep. M. Kasperski.

<sup>12</sup> Ale nie tylko przy procesach fizjologicznych, czy neurofizjologicznych, również po części nie zdajemy sobie sprawę z wszystkich przesłanek w czasie procesów podejmowania decyzji. Stąd w Sztucznej Inteligencji powstanie *data mining*, czyli działu, który zajmuje się wyszukiwaniem wiedzy w wiedzy. Przep. M. Kasperski.

<sup>13</sup> Cechę tą określa się mianem 'plastyczności mózgu'. Przep. M. Kasperski.

<sup>14</sup> W terminologii AI określane jako *parallel distributed processing* [PDP]. Przep. M. Kasperski.

owych procesów lepiej realizowane są przez systemy symboliczne, a inne – przez systemy koneksjonistyczne. Z podanych powyżej argumentów istotnie wynika, że na gruncie koneksjonizmu lepiej modeluje się procesy na niższym poziomie: pamięć, percepcję i motorykę, podczas gdy podejście symboliczne umożliwia realizację wyższych procesów poznawczych: myślenia i rozwiązywania problemów. Druga grupa uważa, że rozwiązaniem problemu: symbolizm czy koneksjonizm? jest integracja obydwu podejść poprzez tworzenie systemów hybrydowych łączących cechy zarówno symboliczne jak i koneksjonistyczne.

#### 4.5. Systemy hybrydowe

Jeśli tylko różne modele są modelami tego samego obiektu, nawet gdy są modelami w różnym sensie słowa ‘model’ (np. model jako zbiór hipotez o zjawisku lub model jako uproszczony system funkcjonujący podobnie jak owo zjawisko), to jest możliwa ich integracja. W przypadku integracji modeli symbolicznych z modelami koneksjonistycznymi proces ten jest bardzo trudny, gdyż choć zakłada się, że modelują one ten sam obiekt – umysł ludzki, to funkcjonują one na różnych poziomach jego opisu. Integracja modeli procesów poznawczych może przebiegać na kilka sposobów:

- zastąpienie [*replacing*]; wszystkie modele na gruncie jednego podejścia zostają zastąpione modelami z drugiego (np. teoria flogistonu zostaje zastąpiona tlenową teorią spalania),
- redukcja [*reduction*]; model na jednym poziomie sprowadzony jest do poziomu niższego, reprezentacje jednego modelu przetłumaczone są na reprezentacje drugiego modelu (np. fizyka Newtona zostaje wyrażona poprzez pojęcia teorii względności),
- subsumcja [*subsumption*]; proces podobny do redukcji, ale różnica polega na zmapowaniu reprezentacji zamiast ich przetłumaczenia (np. zmapowanie instrukcji języka programowania wyższego poziomu na odpowiadające mu instrukcje języka maszynowego),
- połączenie [*merging*]; obydwa modele współwystępują, a każdy z nich wnosi do całości cechy jemu właściwe (Greco 1994).

Systemy hybrydowe są próbą integracji polegającą na założeniu, że obydwa rozważane podejścia dają możliwość wyjaśniania zjawisk poznawczych i są ze sobą kompatybilne. Systemy te dzielą się na dwie grupy. Konstrukcja modeli hybrydowych połączonych fizycznie [*physically hybrid models*] polega na podzieleniu zadania na podzadania, określeniu, które podzadania lepiej wyjaśnia model symboliczny, a które model koneksjonistyczny, zbudowaniu tych modeli i połączeniu [*merging*] ich w jeden system. Przykładem takiego połączenia jest system czytający na głos sekwencje słów: w przypadku słów o regularnej wymowie stosowany jest moduł symboliczny wykorzystujący dobrze zdefiniowane reguły, a w przypadku słów o wymowie nieregularnej – moduł koneksjonistyczny. Moduły rozwiązujące określone podzadania nie są w systemach hybrydowych połączonych fizycznie zbyt ściśle zintegrowane. Modele hybrydowe połączone нефизycznie [*non-physically hybrid models*] tworzone są nie na zasadzie połączenia, ale na zasadzie redukcji lub subsumcji. Pojęcia na gruncie jednego podejścia (najczęściej symbolicznego) wyrażane są w terminach drugiego podejścia (najczęściej koneksjonistycznego). System taki zachowuje cechy obydwu podejść. Przykładem takiego rozwiązania jest system regułowy (system symboliczny na poznawczym poziomie opisu) skonstruowany jako sieć neuropodobna (system koneksjonistyczny na poziomie zbliżonym do poziomu implementacji) (Green i in. 1996, ss. 45-47).

Systemy hybrydowe dają szansę, po pierwsze, na integrację różnych podejść do natury procesów poznawczych, a po drugie, być może na naukowe rozwiązanie problemu psychofizycznego. Dzięki nim dowiadujemy się jak np. strukturalizowaną wiedzę wyrazić w postaci węzłów i połączeń między nimi. Zbliża to naukę do udzielenia odpowiedzi na pytanie: w jaki sposób procesy psychiczne (umysł) pojawiają się w swoim fizycznym substracie (mózgu).

#### 4.6. Ujęcie dynamiczne

W rozdziale o procesach obliczeniowych w nauce poznawczej warto wspomnieć o nie mieszczącym się w założeniach obliczeniowej teorii umysłu nieobliczeniowym podejściu do jego funkcjonowania: podejściu dynamicznym.

W podejściu tym rozpatruje się całość procesów poznawczych jako nieliniowy system dynamiczny opisywany równaniami różniczkowymi. Według dynamicystów procesy poznawcze ulegają nieustannej zmianie w czasie rzeczywistym, a więc podejście obliczeniowe przebiegające w czasie dyskretnym jest nieadekwatne. Podejście dynamiczne zastępuje statyczne reprezentacje atraktorami w przestrzeni fazowej a procesy obliczeniowe na reprezentacjach – dynamiczną zmianą stanu systemu zgodnie z opisującymi go równaniami różniczkowymi (van Gelder i Port 1995).

Powstało wiele modeli dynamicznych modelujących percepcję i motorykę, ale także procesy centralne. Przykładowy system modelujący podejmowanie decyzji (Townsend i Busemeyer 1995) jest opisywany jako układ równań różniczkowych przedstawiających zależności pomiędzy czynnikami wpływającymi na decyzję (np. oczekiwaną korzyścią z danego wyboru). System startuje będąc w pewnym stanie odzwierciedlającym początkowe preferencje i ewoluuje w czasie modelując, według autorów, w zgodzie z danymi psychologicznymi proces podejmowania decyzji. System ten nie odwołuje się jednakże do reprezentacji i obliczeń.

Choć istotnie w obrębie podejścia dynamicznego stworzono sprawnie działające systemy (m.in. roboty), to jego założenia, co do natury procesów poznawczych są wielce kontrowersyjne. Zarzuca się tej teorii wewnętrzną sprzeczność pojęć, pokrywanie się z podejściem koneksjonistycznym (sieci rekurencyjne są przecież nieliniowymi systemami dynamicznymi), niewłaściwą interpretację niektórych pojęć podejścia obliczeniowego (m.in. samego pojęcia 'obliczalność'). Krytyce podlegają też poszczególne rozwiązania, m.in. uważa się, że wspomniana dynamiczna teoria podejmowania decyzji modeluje poprawnie wyłącznie takie decyzje (np. dot. wyboru pokarmu), które znajdujemy już u prostych organizmów. Wnioskowanie na jej podstawie o naturze procesów centralnych człowieka uważa się więc za nieuzasadnione (Eliasmith 1997).

## 5. Reprezentacja

### 5.1. Definicja i rodzaje reprezentacji

Reprezentacją umysłową nazywa się w nauce poznawczej każdy obiekt umysłowy, który pośredniczy między percepcją a zachowaniem się organizmu. W zakresie tego pojęcia zawierają się takie terminy psychologiczne jak myśli, pojęcia, sądy, wyobrażenia, idee itp.

W związku z realizowaną na gruncie nauki poznawczej wspomnianą wcześniej strategią solipsyzmu metodologicznego znaczenie reprezentacji (czyli jej opis na poziomie semantycznym jest wynikiem syntaktycznych relacji do innych reprezentacji. Na podstawie tych relacji dana reprezentacja jest interpretowana przez system poznawczy. Tak więc o sensie danej reprezentacji możemy mówić tylko w odniesieniu do całego systemu reprezentacji [*representational system*], który zastępuje i odzwierciedla pewien zbiór obiektów, a operacje wykonywane w obrębie tegoż systemu odpowiadają dokonywanym na reprezentowanych obiektach.

Ważną cechą reprezentacji, która wynika z ich mediacji pomiędzy bodźcem a reakcją, jest zdolność do bycia przyczyną ludzkiego zachowania. Człowiek nie tylko podejmuje działanie powodowany bodźcem z zewnątrz, ale także w wyniku istnienia określonych reprezentacji w jego umyśle. W teoriach psychologicznych reprezentacje mogą być przyczyną zachowania zarówno w sposób uświadomiony przez człowieka, jak i dla niego nieświadomy (jak ma to miejsce w psychoanalizie) (Greco 1995).

Istnieją dwie podstawowe funkcje reprezentacji. Pierwsza funkcja polega na zastępowaniu obiektów spoza umysłu lub zastępowaniu innych reprezentacji umysłowych. Zastępowanie obiektów spoza umysłu przez reprezentacje jest konieczne, aby móc nimi manipulować i analizować różne ich układy. Manipulacja manualna większością przedmiotów jest bowiem możliwa. Zastępowanie innych reprezentacji umysłowych służy natomiast skróceniu reprezentacji i nadaniu im strukturalizacji. Zamiast przechowywać listę reprezentacji zwierząt latających, upierzonych i posiadających dzióbek zastępujemy ją reprezentacją 'ptak'. W zależności od tego czy reprezentowany jest obiekt (fakt, zjawisko) ze świata fizycznego (np. poprzez zdanie „Jan jest pilotem”) czy reprezentacja umysłowa (np. „Jan myśli, że jest pilotem”) mówimy o reprezentacjach pierwszego lub drugiego rzędu. Druga funkcja reprezentacji polega na tym, że mogą one odzwierciedlać pewną zewnętrzną strukturę. W takim przypadku reprezentacja jest modelem tej sytuacji (np. umysłowy model rozmieszczenia mebli w pomieszczeniu).

Ze względu na dwie powyższe funkcje reprezentacje dzieli się na reprezentacje analityczne/językowe (funkcja zastępowania), które pozostają w arbitralnym, konwencjonalnym związku do obiektu oznaczonego oraz na reprezentacje obrazowe/analogowe, których istotą jest podobieństwo, czyli odzwierciedlanie tego, co reprezentowane. Ze względu na postulowanie istnienia dwóch rodzajów reprezentacji pojawia się problem, w jakiej formie (werbalnej, analogowej lub obu) człowiek przechowuje w pamięci długotrwałej wiedzę (Kurcz 1987, ss. 168-169).

Choć zarówno w systemach sztucznej inteligencji jak i umyśle ludzkim te same obiekty mogą być reprezentowane w różny sposób, to wydaje się zarazem, że człowiek, dzięki ewolucji, zyskał umiejętność reprezentowania różnych rodzajów zjawisk w sposób możliwie najefektywniejszy. Dawałoby to pewną heurystyczną wiedzę, jaką postać reprezentacji zakładać w modelowaniu procesów poznawczych.

Ale można orzekać o efektywności reprezentacji – powinny one spełniać cztery następujące postulaty (Rich i Knight 1992):

- dawać możliwość reprezentowania całej wiedzy w danej domenie,
- pozwalać na tworzenie nowych reprezentacji ze struktur już istniejących,

- tworzenie nowych reprezentacji powinno dążyć w takim kierunku, aby owe reprezentacje najlepiej przybliżyły do rozwiązania danego problemu,
- nowe reprezentacje powinny dawać się łatwo umieszczać w systemie.

Korzystając z pojęcia efektywności niektórzy badacze (np. Pylyshyn) zakładają, że wszelka wiedza pamiętana jest w postaci językowej, gdyż przechowywanie obrazów umysłowych w pamięci jest, wg nich, nieefektywne ze względu na objętość informacyjną i wydłużanie procesu ich przetwarzania. Po wydobyciu z pamięci musiałyby być one powtórnie przetwarzane percepcyjnie w celu analizy (Kurcz 1987, s. 169).

Z kolei teorie podwójnego kodowania (m.in. teoria Paivio) sugerują, że w umyśle istnieją dwa odrębne systemy kodowania reprezentacji, analogowy i werbalny. Wskazują, że występuje między tymi systemami ścisła zależność funkcjonalna: wykazano m.in., że lepiej pamięta się słowa, do których łatwo stworzyć wizualną interpretację. Za teorią podwójnego kodowania przemawia wiele danych psychologicznych, m.in. wspomniane w rozdz. 4.2. badania dotyczące obrazów umysłowych. Potwierdzają tę teorię także wyniki badań neurofizjologicznych: np. w procesie wyobrażania konkretnego przedmiotu w określonym fragmencie kory mózgowej powstaje wzbudzenie elektryczne przypominające kształtem ów przedmiot. Pewne operacje (np. wnioskowanie o przecięciu dwu linii) prowadzone na reprezentacjach kodowanych analitycznie byłyby dużo mniej efektywne od analizy obrazów umysłowych. Z kolei trudno przyjąć analogowe kodowanie struktur językowych. Istnieją więc poważne argumenty na istnienie dwóch [rodzajów] kodów (Thomas 1991).

Istnieją także, choć mało popularne, teorie o więcej niż dwu sposobach reprezentowania. Przykładem może być teoria kodu poszostnego, w której każdemu typowi zmysłów (słuch, wzrok, dotyk, smak, węch oraz osobny kod językowy) odpowiada jeden rodzaj reprezentacji (Tamże).

Oprócz podziału na reprezentacje analogowe i werbalne ważny jest też podział na reprezentacje deklaratywne (wiedza o faktach, wiedza „że”) i reprezentacje proceduralne (wiedza o procesach, wiedza „jak”). Podział ten związany jest nie z formalnym zapisem reprezentacji, ale sposobem interpretacji tego zapisu. Jeśli w danych opisujących pewną reprezentację zawarta jest informacja, jak ją używać, jest to reprezentacja proceduralna, jednak czy dane te zostaną odczytane w taki, a nie inny sposób, zależy od ich interpretacji (Rich i Kight 1992).

Następny podział, a raczej spektrum możliwych rozwiązań, wyznaczane jest przez stopień w jakim dana reprezentacja pozwala systemowi poznawczemu wnioskować o właściwościach semantycznych z jej formalnego zapisu. Począwszy od reprezentacji czysto syntaktycznych, tj. logika formalna, przechodzi się do reprezentacji stworzonych właśnie po to, aby z łatwością uzyskiwać znaczenie danych reprezentacji (sieci semantyczne).

Nawet jeśli reprezentacje kodowane są w jeden sposób, to oczywistym jest fakt, że w świadomości operujemy dwoma rodzajami reprezentacji: językowymi i obrazowymi. W nauce o sztucznej inteligencji wypracowano wiele metod formalnego zapisu obu rodzajów reprezentacji i większość z owych formalnych struktur wykorzystywana jest w nauce poznawczej. Przedstawię zatem najważniejsze rodzaje tych reprezentacji, najpierw rodzinę reprezentacji analitycznych, a następnie analogowych, analizując za każdym razem ich adekwatność w modelowaniu procesów poznawczych.

## 5.2. Reprezentacje analityczne

Najprostszym z analitycznych sposobów reprezentowania wiedzy w systemach sztucznej inteligencji, który da się odnieść do umysłu człowieka jest zapis wiedzy w postaci zdań logiki formalnej: np. rachunku predykatów pierwszego rzędu. W systemie tym predykaty jednoargumentowe reprezentują cechy obiektów (np. wysoki(Jan)), a wieloargumentowe – relacje (np. ojciec(Jan, Tadeusz)). Operacje koniunkcji, alternatywy, równoważności i wynikania opisują zależności między cechami i relacjami, kwantyfikator ogólny pozwala uogólniać cechy, relacje i operacje na pewną klasę obiektów, a szczegółowy – orzekać o istnieniu obiektów o pewnej cesze lub znajdujących się w pewnej relacji.

Reprezentacja logiczna, oprócz niewątpliwej zalety jaką jest jej siła wnioskowania, ma także kilka wad. Już samo wnioskowanie w rachunku logiki obarczone jest wadą tylko częściowej rozstrzygalności. Następnym problemem jest istnienie wielu rodzajów informacji, których nie da się reprezentować poprzez logikę predykatów. Przykładami mogą być zdania zawierające informację niedookreśloną, niepewną, czy heurystyczną. Co więcej, nawet ta możliwa do zapisania w języku logiki informacja nie jest ustrukturalizowana, jest nieuporządkowanym ciągiem formuł atomowych bez żadnej hierarchii. Przykładowe predykaty mające jako argument imię Jana (wysoki(Jan)), ojciec(Jan, Tadeusz), skąpy(Jan), brzydki(Jan) itd.) mogą być jedynymi z wielu formuł znanych systemowi poznawczemu, które znajdują się na tym samym poziomie hierarchii. Wywnioskowanie właściwości semantycznych danego obiektu jest w tym rachunku niezwykle skomplikowane (Rich i Knight 1992, ss. 165-166).

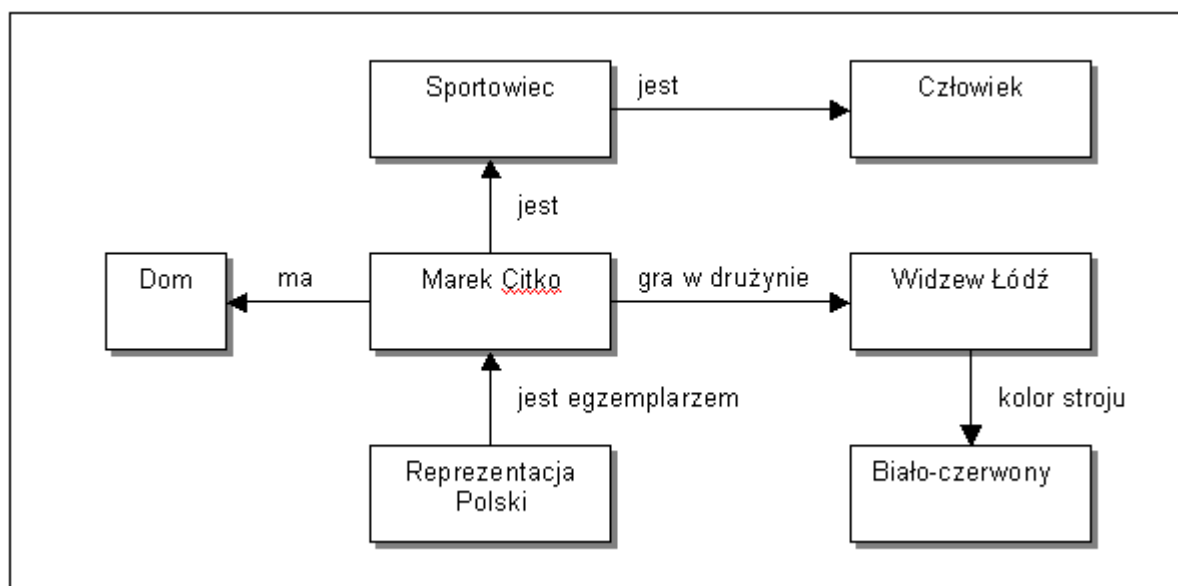
Ścisłe zasady wnioskowania, nastawione na przetwarzanie informacji tak, aby zachować jej prawdziwość, nie są prawdopodobnie metodą używaną przez ludzki umysł. Większość procesów myślenia polega raczej na przeszukiwaniu modeli niż przeprowadzaniu logicznych dedukcji (zob. rozdz. 4.2). Zauważa się jednak, że odpowiednio wytrenowane umysły matematyków i logików są w stanie przyswoić logiczną reprezentację świata (Winograd 1974, s. 65). Odnosi się to zapewne tylko do niektórych rodzajów przetwarzania informacji, tj. dedukcja naturalna, gdyż trudno sobie wyobrazić, by umysł ludzki mógł wnioskować o skomplikowanym problemie stosując np. wspomnianą metodę rezolucji. W tym sensie jednak logika jest narzędziem umysłu a nie jego immanentną cechą, elementem jego struktury a nie biologii.

Dodając do syntaktycznych, logicznych sposobów wnioskowania semantyczną wiedzę o świecie otrzymuje się reprezentację w postaci reguł. Reguły postaci  $A \rightarrow B$ , pozwalają ze zbioru faktów  $A$  wywnioskować zbiór faktów  $B$  lub podjąć akcję  $B$ . Fakty, dzięki inkorporacji w postać reguły, uzyskują niezbędną dla efektywności myślenia strukturalizację. Dzięki przyporządkowaniu każdemu faktowi stopnia jego pewności możliwe jest przetwarzanie wiedzy niepewnej, a dzięki istnieniu w systemie regułowym metareguł – zapis wiedzy heurystycznej. Reguły, traktowane jako wiedza proceduralna, w połączeniu z omawianymi w następnym akapicie reprezentacjami o doskonalszej semantyce pozwalają na tworzenie bogatych modeli systemu poznawczego, takich jak opisane w rozdz. 4.2 systemy ACT\* i SOAR.

W nauce o sztucznej inteligencji stworzono także rodzinę reprezentacji, które umożliwiają w dogodny sposób modelowanie własności semantycznych obiektów. Przechowują własności danego obiektu w jednej strukturze oraz pozwalają na dziedziczenie własności klasy przez obiekty do niej należące. Reprezentacje te dzielą się na dwie grupy. Pierwsza grupa to ogólniejsze modele, umożliwiające zapisanie wiedzy z wielu dziedzin, czyli reprezentację różnych klas obiektów i ich systemów. Nie wykorzystują one jednak specyfiki konkretnej

dziedziny do optymalizacji przetwarzania wiedzy z tej dziedziny. Podstawowe typy takich struktur to sieci semantyczne oraz ramy<sup>15</sup>. Druga grupa to reprezentacje, których struktura uwzględnia wiedzę o cechach informacji, jaką mają one przechowywać (Rich i Knight 1992, s. 251-252). Wyróżnić tu można m.in. popularne w psychologii poznawczej schematy i prototypy, oraz wywodzące się z nauki o sztucznej inteligencji skrypty (Kurcz 1987, 1961-1962).

Głównym założeniem, które przyjęli twórcy sieci semantycznych (m.in. Quillian) jest pogląd, że znaczenie danego pojęcia pochodzi od jego relacji do innych pojęć. Sieć semantyczna jest zatem zbiorem pojęć połączonych relacjami. Relacje te mogą być różne: od najbardziej podstawowych, jak np. „jest” lub „jest egzemplarzem”, po relacje specyficzne dla konkretnej dziedziny (np. sportu „gra w drużynie”, „kolor stroju” itp.) (Rich i Knight, rozdz. 91.1).



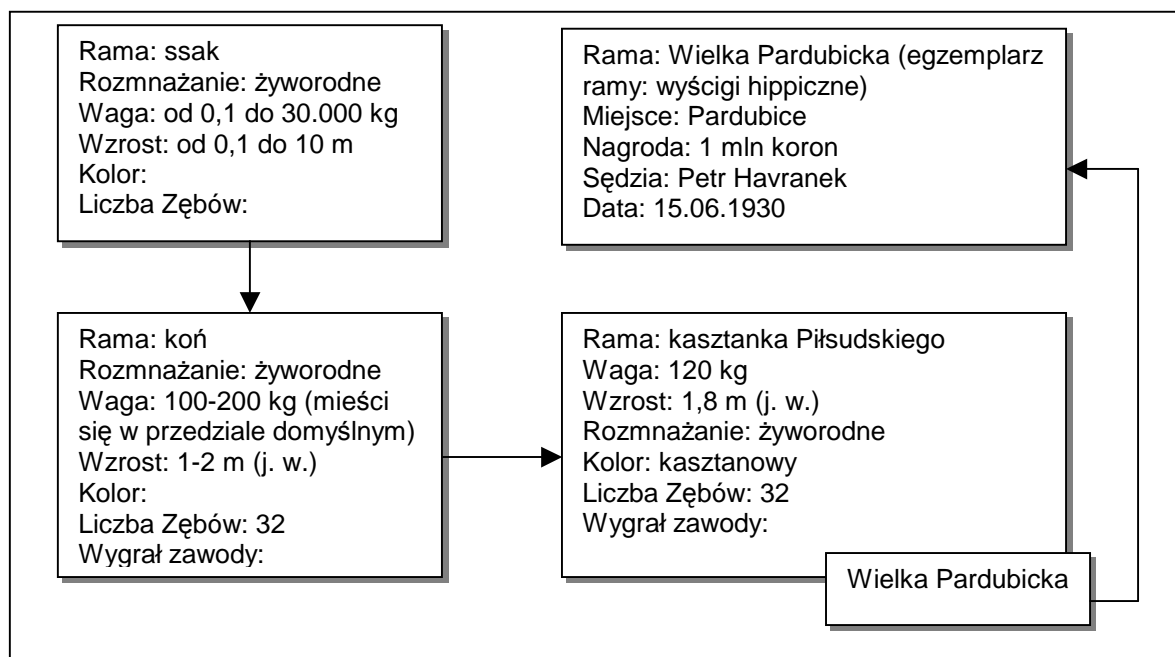
Rysunek 7. Przykład sieci semantycznej (oprac. własne).

Z powyższego schematu możemy wywnioskować, że Marek Citko jest sportowcem, ma dom i jest egzemplarzem reprezentacji Polski (rozumianej kolektywnie). Proces prowadzący do odpowiedzi na pytania na podstawie sieci semantycznej zwany jest ‘poszukiwaniem przecięcia’ [*intersection search*]. W eksperymentach psychologicznych potwierdzono, że sieć semantyczna może być dobrym modelem organizacji pojęć w umyśle. Wskazano m.in., że czym więcej węzłów w sieci semantycznej dzieli od siebie dwa pojęcia, tym dłużej człowiek odpowiada na pytanie o relację między nimi. Jeśli więc badany miałby pojęcie dotyczące Marka Citki zorganizowane w przedstawiony powyżej sposób, prawdopodobnie krócej zastanawiałby się nad odpowiedzią na pytanie: „w jakiej drużynie gra Marek Citko?” niż na pytanie „jaki jest kolor jego stroju?”. Jednak czym bardziej skomplikowane pytania, tym bardziej skomplikowana musi być struktura sieci reprezentującej wiedzę niezbędną do odpowiedzi. Dodatkowo, niezbędne staje się strukturalizowanie pojęć, początkowo oznaczanych tylko nazwą. Chociaż granica ta jest płynna, czym bardziej ustrukturalizowana jest sieć semantyczna, tym bardziej nabiera cech systemu ram.

<sup>15</sup> Od ang. *frames*. Przyp. M. Kasperski.



Według Marvinina Minsky'ego (Minsky 1975), twórcy teorii ram [*frames*], człowiek wykorzystuje rami, gdy napotyka nową sytuację (problem) lub dotychczasowy problem zostaje postrzeżony z innego punktu widzenia. Rama jest strukturą składającą się z klatek [*slots*] zawierających różne rodzaje informacji o jednym konkretnym obiekcie lub sytuacji. Niektóre klatki reprezentują wiedzę proceduralną o tym, jak używać daną ramę, inne zawierają konkretne wartości danego atrybutu, wartości domyślne lub wartości spodziewane. Niektóre klatki mogą zawierać też inne rami. Zbiór połączonych ze sobą ram tworzy system [*frame system*]. Za relacje pomiędzy ramami odpowiadają pola zawierające wskaźniki do innych ram, odpowiadające takim zależnościom jak: bycie egzemplarzem klasy, bycie częścią obiektu i bycie podklasą klasy. Taka hierarchia umożliwia dziedziczenie własności przez rami np. egzemplarz danego pojęcia dziedziczy wszystkie wartości atrybutów definicyjnych tego pojęcia, ale zawiera swoiste wartości dla atrybutów akcydentalnych. Rama-koń dziedziczy zatem właściwość żyworodności właściwą ramie-prototypowi-ssak, ale ma indywidualny wzrost i wagę. Ma także nową klatkę: „wygrane zawody”, o określonej zawartości tylko dla konkretnego konia. Dziedziczenie nie zachodzi dla elementów danego zbioru rozumianych kolektywnie. Teoria ram [pozwala z jednej strony na naturalne i elastyczne reprezentowanie obiektów, a z drugiej – daje możliwość wnioskowania i uogólniania wiedzy.



Rysunek 8. Przykładowa hierarchia ram: rama prototyp-ssak, jej egzemplarz rama-koń i jednocześnie prototyp dla rami-kasztanka. Rama-kasztanka zawiera ramę-Wielka Pardubicka.

Rami, podobnie zresztą jak sieci semantyczne, są ogólnym modelem reprezentacji, tak ogólnym, że trudno empirycznie weryfikowalnym i zarazem nieefektywnym obliczeniowo. Implementacje teorii ram do konkretnej dziedziny wykorzystują wiedzę o tej dziedzinie do optymalizacji tworzonych struktur i metod wnioskowania na ich podstawie. Jedną z takich szczegółowych wersji teorii ram jest koncepcja skryptów [*scripts*] Schanka i Abelsona (Schank i Abelson 1975). Skrypty to reprezentacje na bazie ram zawierające informację o sekwencji zdarzeń (czynności) zachodzących w określonym kontekście. Przechowują one wiedzę podmiotu, jak zachować się w standardowej, dobrze znanej sytuacji. Dany skrypt nie

ulega znaczącym zmianom, nie daje także możliwości radzenia sobie z sytuacjami nowymi i niestandardowymi. Do tego celu wspomniani autorzy stworzyli koncepcję „planu”. Strukturę i zawartość skryptu, czyli układ jego klatek, opisują pewne podstawowe jednostki semantyczne zaproponowane przez Schanka w ramach teorii zależności pojęciowej [*conceptual dependency*]. Są to najbardziej podstawowe typy czynności (np. PTRANS – fizyczne przemieszczenie się obiektu; ATRANS – przemieszczenie abstrakcyjnej relacji, np. posiadania; MOVE – przemieszczenia ciała lub jego części czy MBUILD – utworzenie nowej informacji) i obiektów (np. PP – *picture producer*; czyli obiekt fizyczny odbierany zmysłem wzroku). Za pomocą owych jednostek opisywana jest standardowa sytuacja. Przykładowy skrypt opisujący wizytę w restauracji wygląda następująco:

<b>Nazwa skryptu:</b> restauracja	<b>Osoby:</b> klient, kelner, kucharz, kasjer
<b>Cel:</b> uzyskanie jedzenia, aby zaspokoić głód i uzyskać przyjemność	
<b>Scena 1:</b> wejście PTRANS ja do restauracji ATTEND oczy, gdzie wolne miejsce MBUILD gdzie usiąść PTRANS ja do stolika MOVE zajęcie miejsca	<b>Scena 2:</b> zamówienie ATrans otrzymanie karty MTRANS przeczytanie karty .... <b>Scena 3:</b> jedzenie .... <b>Scena 4:</b> wyjście .... PTRANS ja z restauracji

Rysunek 9. Przykładowy skrypt: wizyta w restauracji (na podstawie Schank i Abelson 1975, s. 424).

Kolejne czynności są ze sobą powiązane, czynność następna może zajść dopiero po spełnieniu poprzedniej. Umożliwia to uzależnienie reakcji podmiotu od zachodzących zdarzeń, szczególnie od zdarzeń odbiegających od prawidłowego scenariusza. Klient może np. nie zapłacić napiwku, jeśli podane mu jedzenie było zimne i nie może zjeść jedzenia, które nie zostało mu jeszcze przyniesione. Za to, który skrypt zostanie przywołany przez system poznawczy odpowiada jego nagłówek: nazwa, osoby i cel. Nagłówek ten musi pasować do obecnej sytuacji – inny skrypt zostanie uruchomiony, jeśli wchodzimy do restauracji, aby zarezerwować stolik na weekend.

Wiele eksperymentów potwierdziło istnienie reprezentacji umysłowych o cechach skryptów: reprezentacji standardowych sytuacji z życia codziennego charakteryzujących się automatyzmem zachowania podmiotu. Np. eksperyment E. Langer (Bobyk 1987, s. 64) polegał na tym, że do pierwszej z osób czekających w kolejce na skorzystanie z kserokopiarki podchodziła inna osoba, prosząc o przepuszczenie i motywując swoją prośbę w różny sposób: nadzwyczajnym pośpiechem, informacją placebo („muszę zrobić te odbitki”) lub w ogóle jej nie uzasadniając. Okazało się, że gdy chodziło o kilka kopii, to każde uzasadnienie było równie efektywne, proszony uruchamiał bowiem wg eksperymentatora skrypt pt. „drobna uprzejmość”. Przy dużej liczbie kopii proszony nie działał już automatycznie i uzasadnienie zyskiwało znaczenie dla podjęcia przez niego decyzji.

Prowadzone są także badania mające na celu potwierdzenie leżącej u podłoża teorii skryptów hipotezy o istnieniu pierwotnych elementów (jednostek) znaczeniowych. Do najsłynniejszych należą prace Anny Wierzbickiej, która w wyniku analizy wielu języków wyodrębniła pojęcia występujące we wszystkich z nich i nie dające się dalej uprościć. Okazało się, że zbiór tych pojęć jest niewielki, docelowo przewiduje się kilkadziesiąt jego elementów. M.in. wyodrębniono pięć uniwersalnych pojęć (zwanymi indefinibiliami) dotyczących czynności umysłowych: myśleć, mówić, wiedzieć, czuć i chcieć (Wierzbicka 1995, s. 236).

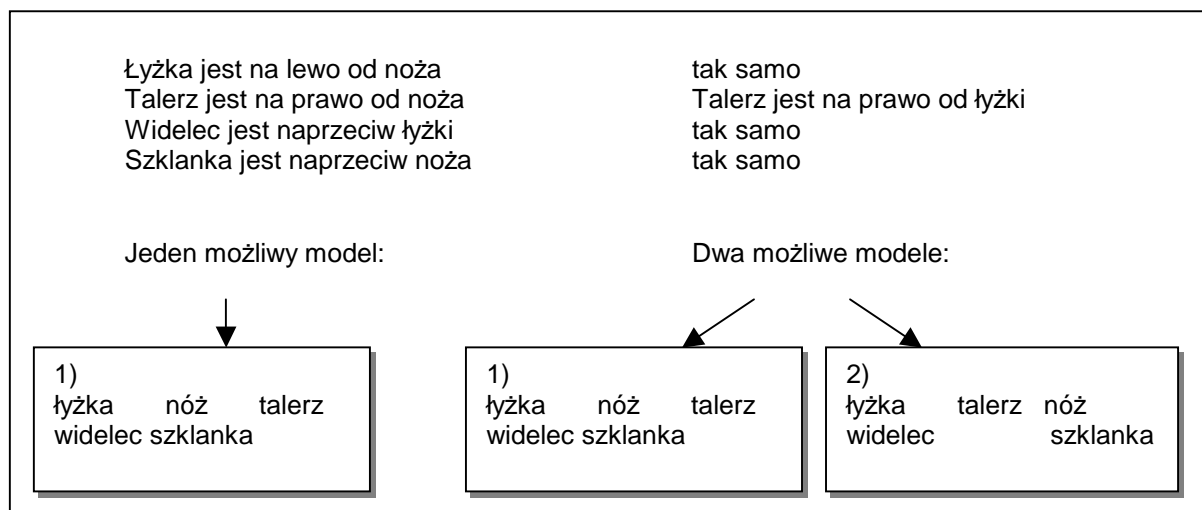
### 5.3. Reprezentacje odzwierciedlające

Drugim z rodzajów reprezentacji umysłowych są reprezentacje odzwierciedlające. Przykładem takich reprezentacji mogą być, mające doniosłe znaczenie w nauce poznawczej, modele umysłowe. Choć ich istnienie postulowano już w latach czterdziestych, to teoria modeli została rozwinięta w pełni w latach osiemdziesiątych, m.in. przez Philipa Johnsona-Lairda (Johnson-Laird 1989). Według niego umysł człowieka konstruuje modele o strukturze podobnej do struktury zewnętrznej sytuacji i dokonuje na tych modelach takich operacji, które imitują operacje dokonywane w świecie fizycznym. Wynik operacji ma skutek behawioralny powodując konkretne zachowanie człowieka i poznawczy – człowiek może analizować zgodność stworzonego modelu ze światem zewnętrznym.

Modele stosowane są wg Johnsona-Lairda podczas wszystkich procesów poznawczych. Percepcja stanowi źródło modeli, człowiek scala informacje płynące ze wszystkich zmysłów (choć głównie informacje wzrokowe) w jeden model przestrzenny. Informacje płynące z rozmowy, choć mają postać zdań nie odzwierciedlających struktury sytuacji, powodują powstanie w umyśle jej modelu. Człowiek o tyle rozumie czyjąś wypowiedź (ustną lub pisemną), o ile potrafi skonstruować do niej odpowiedni (jeden lub więcej) model umysłowy. Człowiek uzna wypowiedź za prawdziwą, jeśli uda mu się wpasować choć jeden model odpowiadający znaczeniu wypowiedzi w jego model świata. Skomplikowanie modelu świata o możliwościach poznawczych jego posiadacza: „granice naszych modeli to granice naszego świata” – pisze Johnson-Laird parafrazując Wittgensteina.

Za pomocą modeli człowiek przeprowadza myślenie. Wpierw konstruuje model na podstawie przesłanek – danych początkowych. Ujęcie tych danych w postać modelu ujawnia zazwyczaj pewną nieuświadomioną wcześniej zależność pomiędzy danymi – konkluzję. Umysł szuka następnie modeli sprzecznych z uzyskaną konkluzją, ale zgodnych z przesłankami i modelami świata. Jeśli modele takie istnieją, konkluzja zostaje odrzucona, jeśli zaś nie istnieją, to zostaje utworzony nowy model lub modele zawierające nową informację. Tak opisany proces myślenia to ujęcie procesów rozumowania jako przeszukiwania przestrzeni stanów.

Johnson-Laird podaje wiele badań potwierdzających istnienie modeli umysłowych i ich wykorzystywanie przez umysł zgodnie z postulatami teorii. Dwa najważniejsze przykłady to eksperymenty dotyczące rozumienia zdań i wnioskowania sylogistycznego. Pierwszy eksperyment polegał na przedstawieniu badanym par zbiorów zdań, z których jeden był jednoznacznie interpretowalny, a drugi powodował powstanie kilku konkurencyjnych modeli. Badane osoby dużo lepiej pamiętały szczegóły sytuacji o jednoznacznej interpretacji niż o interpretacji wieloznacznej.



Rysunek 10. Możliwe interpretacje dwóch zbiorów zdań (na podstawie: Johnson-Laird 1989, s. 472).

Teoria modeli umysłowych dobrze wyjaśniała i przewidywała procesy badane w drugim eksperymencie (a właściwie ich serii). Otóż od dawna było wiadomo, że spośród 256 możliwych sylogizmów człowiek rozpoznaje niektóre jako poprawne bardzo łatwo (np. w wieku już 9 lat), a inne są rozpoznawane tylko przez niewielki procent badanych. Podobnie jest z umiejętnością wywnioskowania konkluzji sylogizmu z dwóch jego przesłanek. Badania dowiodły, że stopień trudności sylogizmu jest proporcjonalny do skomplikowania i liczby modeli umysłowych możliwych do konstrukcji na jego podstawie, teoria modeli tłumaczy także błędy popełniane podczas wnioskowania (Johnson-Laird 1993, ss. 32-36).

Istotą modeli jest odzwierciedlenie struktury sytuacji, niekoniecznie elementy modelu muszą przypominać reprezentowane obiekty. Mogą one mieć postać obiektów pochodzących z percepcji lub wyobraźni, ale alternatywnie mogą mieć postać symboli [*symbol tokens*], reprezentujących nie tylko obiekty fizyczne, ale także abstrakcyjne relacje (negację, przynależność itp.). Wspomniane wcześniej obrazy umysłowe traktuje się jako specjalny rodzaj modeli, dwuwymiarową reprezentację geometryczną powstałą z trójwymiarowego modelu (Johnson-Laird 1989, ss. 488-490).

#### 5.4. Reprezentacje koneksjonistyczne

Trzecim rodzajem reprezentacji, o którym trudno orzec czy symbolizuje czy też odzwierciedla obiekty świata zewnętrznego jest reprezentacja na poziomie subsymbolicznym – reprezentacja koneksjonistyczna. O ile w przypadku wersji lokalistycznej, gdzie jeden węzeł odpowiada jednemu obiektowi czy pojęciu, możemy się pokusić o analogię do symbolu, to w przypadku reprezentacji rozproszonej sytuacja jest o wiele bardziej skomplikowana. Stan aktywacji określony na zbiorze jednostek przetwarzających w przypadku sieci modelujących percepcję istotnie może przypominać obraz przetwarzanego obiektu, trudno jednak stosować termin ‘podobieństwo’ w stosunku np. do reprezentacji reguł gramatycznych w sieciach rozpoznających poprawność zdań, sekwencji ruchów w sieci Jordana itp. Poziom subsymboliczny jest bowiem zbyt zbliżony do poziomu implementacji, aby pojęcia ‘symbolizowanie’ czy ‘podobieństwo’ miały sens, podobnie jak nie ma sensu analizowanie

struktury danych w pamięci komputera w postaci ładunków elektrycznych w jego kościach pamięci.

Koneksjonistyczna reprezentacja rozproszona ma istotne cechy. Po pierwsze jest to struktura bardzo ogólna (podobnie jak ramy Minsky'ego), która dopiero w konkretnej implementacji zyskuje uszczegółowienie. Uzyskuje ona interpretację dopiero na podstawie analizy stanu aktywacji określonego na zbiorze wszystkich jednostek przetwarzających (reprezentacja odpowiadająca pamięci krótkotrwałej) oraz sile i strukturze połączeń pomiędzy nimi (reprezentacja odpowiadająca pamięci długotrwałej). Czasami, szczególnie w przypadku warstw ukrytych, w ogóle taka interpretacja jest niemożliwa. reprezentacje podobnych pojęć czy obiektów nakładają się na siebie (w tym sensie sieć odzwierciedla podobieństwo), są także odporne na usunięcie czy zniszczenie swoich fragmentów<sup>16</sup>. Reprezentacja ta ma także dobre potwierdzenie w badaniach neurofizjologicznych. M.in. w badaniach na szczurach nauczonych przechodzić labirynt nie stwierdzono zależności między obszarem mózgu, z którego wycinano tkankę nerwową a utratą nabytej zdolności, ale stwierdzono taką zależność między ilością wyciętej tkanki. Potwierdzałoby to duży stopień rozproszenia reprezentacji w mózgu. Potwierdzenie napływa także z psychologii, np. wyniki badań psychologicznych dotyczących zapominania dobrze wyjaśniane są przez nakładanie się podobnej reprezentacji na drugą, co powoduje zacieranie reprezentacji wcześniejszej (Rich i Knight 1992, ss. 520-522).

## Zakończenie

Opisując osiągnięcia nauki poznawczej w wyjaśnianiu niektórych procesów poznawczych należy zdawać sobie sprawę, jak niewiele procesów zostało poznanych i zrozumianych w porównaniu z bogactwem funkcji umysłu. Młoda nauka, badając niezwykle skomplikowane zjawisko, musiała z konieczności ograniczonych zasobów – czasu, liczby naukowców, funduszy, a przede wszystkim dotychczas zgromadzonej wiedzy – rozpocząć od relatywnie prostszych elementów zagadki: percepcji, pamięci i wnioskowania. Procesy powyższe najlepiej mieszczą się bowiem w założeniach obliczeniowej teorii umysłu i najłatwiej poddają się formalizacji.

Jeśli jednak teoria ta jest słuszna, modelowaniu komputerowemu powinny się poddać także inne zjawiska umysłowe: procesy emocjonalne, twórczość, zachowania społeczne, marzenia senna, zachowania człowieka na polu etyki i estetyki. Obecnie są to prawie dziewicze przedmioty badań, o których poprzez nieliczne istniejące modele niewiele się na razie dowiedziano. Jako przykład pracy w tych poddziedzinach można wymienić program Daydreamer symulujący powstawanie wyobrażeń (m.in. marzeń), w celu przypominania lub przewidywania określonych sytuacji, regulacji emocji oraz wsparcia dla procesów twórczych. Program generuje opisy wyobrażeń na podstawie zawartości pamięci epizodycznej, wiedzy ogólnej, celów i stanów emocjonalnych (Mueller i Dyer 1985). Innym przykładem jest ACRES, program modelujący emocje. Jego autorzy, wśród wielu postulatów dotyczących przebiegu ludzkich emocji, zauważają, że emocje można modelować tylko obok innych funkcji poznawczych, gdyż trudno sobie wyobrazić organizm, który nie robi nic poza przeżywaniem emocji (Moffat i in.). Powstał także program dokonujący ocen estetycznych zdjęć twarzy kobiet na podstawie stopnia prostoty, z jaką potrafi on zakodować, po

<sup>16</sup> Co zbliża ją tym samym do cechy mózgu naturalnego, która zwana jest 'plastycznością mózgu'. Przyp. M. Kasperski.

określonych przekształceniach geometrycznych, zadane zdjęcie. Przedstawiono porównanie ocen programu z sędziami ludzimi oraz regułami, jakie stosowali słynni malarze przy tworzeniu portretów (Schmidhuber 1998). Oprócz rzeczywiście działających programów istnieje także w wymienionych poddziedzinach wiele prac teoretycznych, które na razie nie znalazły realizacji postaci modeli komputerowych.

Poza sferą osiągnięć nauki poznawczej pozostaje także najtrudniejsze chyba do wyjaśnienia pytanie związane z umysłem: zagadnienie świadomości. Pomimo istnienia wielu poglądów na temat zjawiska świadomości, nie istnieje ani teoria akceptowana przez większość badaczy, ani taka, którą dałoby się zaimplementować w modelu. Zjawisko świadomości pozostaje wciąż jednym z największych wyzwań nauki, które prawdopodobnie nieprędko zostanie wyjaśnione. Krytyka nauki poznawczej nie ogranicza się tylko do wskazywania tych zjawisk psychicznych, o których naukowcy poznawczy mają niewiele do powiedzenia. Wielu naukowców i filozofów kontestuje podstawy obliczeniowej teorii umysłu. Sądzą oni, że wykorzystanie metod sztucznej inteligencji do badania umysłu człowieka opiera się na błędnych założeniach, że jest to droga, która nie prowadzi do wyjaśnienia zachowania się człowieka. Krytyków tych można najogólniej podzielić na dwie grupy. Przedstawiciele pierwszej uważają, że umysł należy badać nie na poziomie operacji formalnych, ale na poziomie wyższym. Twierdzą, że wielu procesów psychicznych nie da się sprowadzić do operacji na symbolach czy sieciach aktywacji. Uważają, że są to procesy (jak np. zjawiska świadomości czy rozumienia) samoistne, niezależne od struktur symboli znajdujących się w umyśle. Badacze należący do drugiej grupy mają zdanie przeciwne. Sądzą oni, że poziom syntaktyczny to poziom za wysoki do opisu zachowania człowieka. Wg nich, naukowego wyjaśnienia dostarczyć może jedynie badanie układu nerwowego. Oczywiście podział powyższy jest bardzo upraszczający, gdyż w każdej z dwóch grup istnieje szerokie spektrum poglądów: od radykalnych, np. kwestionujących w ogóle istnienie umysłu jako pojęcia naukowego (materializm eliminujący<sup>17</sup>), aż do poglądów umożliwiających, wraz z rozwojem nauki, integrację z nauką poznawczą, lub nawet redukcją czy to nauki poznawczej, czy też owych nurtów.

Obecnie trudno jest rozstrzygnąć, czy rację mają krytycy, czy też zwolennicy stosowania metod sztucznej inteligencji w badaniu umysłu. Dotychczas nie odnotowano na tyle spektakularnych osiągnięć, aby pozwoliły one na empiryczne obalenie też krytyków. Z drugiej strony, nierozstrzygnięta jest też sama krytyka. Neurofizjologia, mimo swego ogromnego postępu, ma przed sobą jeszcze wiele nieodkrytych tajemnic. Jak na razie, daleko do wyjaśnienia jak myśl powstała w naszym umyśle (np. „niniejsza praca jest ciekawa”) wyjaśnić zajściem określonych procesów w mózgu (np. uaktywniły się neurony nr 514 i 1022). Prace filozofów i psychologów krytykujących naukę poznawczą same zostały gruntownie skrytykowane, a niektórzy spośród krytyków przyznają, że jeśli nie wszystkie, to przynajmniej część procesów psychicznych człowieka jest upośredniona symbolicznie. Jako takie mogą być one sensownie badane przy zastosowaniu metod sztucznej inteligencji (zob. Bobryk 1992, ss. 75-76).

Zastosowanie metod sztucznej inteligencji w badaniach umysłu, czyli rozwój nauki poznawczej, niesie za sobą tak obiecujące perspektywy, że jest to dziedzina warta największej uwagi. Obliczeniowa teoria umysłu sama w sobie jest, podobnie jak wynalazek komputera, jednym z największych osiągnięć naszej cywilizacji. Dzięki niezwykłemu sprzężeniu zwrotnemu dwa podstawowe filary nauki poznawczej: psychologia i nauka o sztucznej

<sup>17</sup> Czyli, po prostu *eliminatywizm*. Przep. M. Kasperski.

inteligencji zyskały niezwykle dynamikę rozwoju i bogate osiągnięcia. Korzyść psychologów polega na uzyskaniu ścisłej metody naukowej o dużym stopniu obiektywizmu. Pozwala ona nie tylko przewidywać zachowanie człowieka, ale także konstruować i weryfikować hipotezy na temat wewnętrznych, nieobserwowalnych zjawisk (stanów umysłu) prowadzących do tego zachowania. Psychologia dowiadyuje się nie tylko „jak?”, ale i „dlaczego?”. Dominacja nurtu poznawczego we współczesnej psychologii nie jest więc zaskoczeniem. Z kolei nie sposób budować efektywne maszyny obdarzone inteligencją nie poznając jedyne naturalnego źródła inteligencji – ludzkiego umysłu. Dzięki badaniom nauki poznawczej mogło powstać wiele systemów sztucznej inteligencji, których, bez poznania odpowiednich procesów poznawczych człowieka, stworzyć nie byłibyśmy w stanie. Systemy te już dzisiaj wykorzystywane są w gospodarce, a wraz ze wzrostem udziału gałęzi opartych na przetwarzaniu informacji w PKB krajów rozwiniętych, wykorzystanie to stanie się zapewne jeszcze powszechniejsze. Ogromne są także możliwości praktycznego wykorzystania gromadzonej wiedzy o tym, jak człowiek poznaje świat: optymalizacja edukacji, przekazów informacyjnych (m.in. reklamy), psychoterapia poznawcza, lepsze wykorzystanie możliwości umysłów ludzkich (np. umysłów naukowców).

Dwadzieścia pięć lat rozwoju badań nad tak trudnym zjawiskiem, jakim jest funkcjonowanie umysłu, to z pewnością zbyt krótki okres, aby ferować dalece idące wnioski i przewidywać przyszłe osiągnięcia. Dwadzieścia pięć lat to średnia długość kariery naukowej – większość twórców najbardziej uznanych teorii nauki poznawczej brała udział w tworzeniu tej dziedziny. Być może młode pokolenia wniosą wkrótce pomysły rewolucjonizujące badania nad umysłem. Ukoronowaniem tych badań, potwierdzającym, że o umyśle wiemy już naprawdę dużo, byłaby konstrukcja maszyny zdolnej do przejścia testu Turinga. Myślę, że zarówno nieuprawnione jest podawanie konkretnych dat takiego wydarzenia, jak i przeciwnie – dowodzenie, że budowa sztucznej inteligencji o wszystkich cechach inteligencji naturalnej jest z natury niemożliwa. Być może słuszne jest twierdzenie filozofa i futurologa, Stanisława Lema, który w eseju *Tajemnica chińskiego pokoju* (Lem 1996, s. 207) napisał, iż to, czy sztuczna inteligencja jest możliwa, może pokazać jedynie przyszłość.

Wiosna/jesień 1999

### Literatura:

- [1] J. Bobryk, *Locus umysłu*, Zakład Narodowy im. Ossolińskich, Wrocław 1987.
- [2] J. Bobryk, *Niektóre problemy epistemologiczne w badaniach 'cognitive science'*, w: *Pogranicza epistemologii*, red. J. Niżnik, IFiS PAN, Warszawa 1992.
- [3] J. Bobryk, *Akty świadomości i procesy poznawcze*, Zakład Narodowy im. Ossolińskich, Wrocław 1994.
- [4] L. Borkowski, *Wprowadzenie do logiki i teorii mnogości*, Towarzystwo Naukowe KUL, Lublin 1991.
- [5] G. Bower, J. Clapper, *Experimental Methods in Cognitive Science*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [6] D. Chalmers, *A Computational Foundation for the Study of Cognition*, Internet 1997.

- [7] Z. Chlewiński, *Umysł. Dynamiczna organizacja pojęć*, Wydawnictwo Naukowe PWN, Warszawa 1999.
- [8] F. Crick, *Zdumiewająca hipoteza*, Prószyński i S-ka, Warszawa 1997.
- [9] D. Dennett, *The Intentional Stance*, MIT Press, Cambridge, MA 1987.
- [10] K. Domańska, *Metafora komputerowa w psychologii poznawczej*, w: "Prakseologia", Nr 1-2/1991.
- [11] C. Eliasmith, *Computational and Dynamical Models of Mind*, w: "Mind and Machines", Nr 7/1997, ss. 531-541.
- [12] M. Eysenck (red.), *The Blackwell Dictionary of Cognitive Psychology*, Basil Blackwell, Cambridge, MA 1990.
- [13] J. Fodor, *The Language of Thought*, The Harvester Press, Hassocks, Sussex, England 1976.
- [14] J. Fodor, *Methodological solipsism considered as a research strategy in cognitive psychology*, w: "Behavioral and Brain Sciences", Nr 3/1980, ss. 63-109.
- [15] T. van Gelder, R. Port, *It's About Time: An Overview of the Dynamical Approach to Cognition*, w: *Minds as Motion*, red. R. Port, T. van Gelder, Bradford Books/MIT Press, Cambridge, MA 1995.
- [16] A. Greco, *Intergrating 'Different' Models in Cognitive Psychology*, w: "Cognitive Systems", Nr 1(4)/1994, ss. 21-32.
- [17] A. Greco, *The Concept of Representation in Psychology*, w: "Cognitive Systems", Nr 2(4)/1995, ss. 247-256.
- [18] D. Green i inni (red.), *Cognitive Science. An Introduction*, Blackwell, Cambridge, MA 1996.
- [19] R. Gregory, *The Oxford Companion to The Mind*, Oxford University Press, Oxford 1987.
- [20] P. Johnson-Laird, *The Computer and the Mind*, Harvard University Press, Cambridge, MA 1988.
- [21] P. Johnson-Laird, *Mental Models*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [22] P. Johnson-Laird, *Human and Machine Thinking*, Lawrence Erlbaum, Hillsdale, NJ 1993.
- [23] M. Jordon, D. Rosenbaum, *Action*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [24] J. Koziński, *Koncepcje psychologiczne człowieka*, Żak, Warszawa 1996.
- [25] W. Krajewski, *Prawa nauki*, Książka i Wiedza, Warszawa 1998.
- [26] I. Kurcz, *Język a reprezentacja świata w umyśle*, PWN, Warszawa 1987.
- [27] S. Lem, *Tajemnica Chińskiego Pokoju*, TAIWPN Universitas, Kraków 1996.
- [27] J. Lyons, *Chomsky*, Prószyński i S-ka, Warszawa 1998.
- [28] W. Lyons, *Wstęp do Modern Philosophy of Mind*, Everyman, Londyn i Vermont 1995.
- [29] T. Maruszewski, *Psychologia poznawcza*, Polskie Towarzystwo Semiotyczne, Warszawa 1996.



- [30] G. Miller, *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information*, w: "The Psychological Review", Nr (63)/1956, ss. 81-97.
- [31] M. Minsky, *Frame-System theory*, w: *Thinking: Readings in Cognitive Science*, red. P. Johnson, P. Wason, Cambridge University Press, Cambridge 1975.
- [32] D. Moffat, N. Frija, H. Phaf, *Analysis of a model of emotions*, Internet.
- [33] E. Mueller, M. Dyer, *Towards a computational theory of human daydreaming*, w: *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, Irvine, CA 1985, ss. 120-129.
- [34] J. Mulawka, *Systemy ekspertowe*, WNT, Warszawa 1996.
- [35] R. Murawski, *Filozofia matematyki*, Wydawnictwo Naukowe PWN, Warszawa 1995.
- [36] A. Newell, H. Simon, *GPS, A Program That Simulates Human Thought*, w: *Computers and Thought*, red. E. Feigenbaum, J. Feldman, MIT Press, Cambridge, MA 1961.
- [37] A. Newell, P. Rosenbloom, J. Laird, *Symbolic Architectures for Cognition*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [38] A. Newell, R. Young, T. Polk, *The Approach Through Symbols*, w: *The Simulation of Human Intelligence*, red. D. Broadbent, Basil Blackwell, Oxford 1993.
- [39] S. Osowski, *Sieci neuronowe w ujęciu algorytmicznym*, WNT, Warszawa 1996.
- [40] R. Penrose, *Nowy umysł cesarza*, Wydawnictwo Naukowe PWN, Warszawa 1996.
- [41] R. Piłat, *Teorie semantyczne a informatyczne podejście w epistemologii*, w: *Pogranicza epistemologii*, red. J. Niżnik, IFiS PAN, Warszawa 1992.
- [42] S. Pinker, *Language Acquisition*, w: *Foundations of Cognitive Science*, Bradford Books/MIT Press, Cambridge, MA 1989.
- [43] V. Pratt, *Thinking Machines*, Basil Blackwell, Oxford 1987.
- [44] H. Putnam, *Mind, language and Reality*, Cambridge University Press, Cambridge 1975.
- [45] Z. Pylyshyn, *Computing in Cognitive Science*, w: *Foundations of Cognitive Science*, Bradford Books/MIT Press, Cambridge, MA 1989.
- [46] E. Rich, K. Knight, *Artificial Intelligence*, McGraw-Hill, Nowy Jork 1991.
- [47] C. Rollins, *Solipsism*, w: *The Encyclopedia of Philosophy*, red. P. Edwards, Nowy Jork 1967.
- [48] D. Rumelhart, *The Architecture of Mind: A Connectionist Approach*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [49] P. Salo, *What Is Cognitive Science?*, Internet 1997.
- [50] R. Schank, R. Abelson, *Scripts, plans and knowledge*, w: *Thinking: Readings in Cognitive Science*, red. P. Johnson, P. Wason, Cambridge University Press, Cambridge 1975.
- [51] J. Schmidhuber, *Facial Beauty and Fractal Geometry*, Note IDSIA-28/1998.
- [52] S. Scribner, *Modes of Thinking and Ways of Speaking: Culture and Logic Reconsidered*, w: *Thinking: Readings in Cognitive Science*, red. P. Johnson, P. Wason, Cambridge University Press, Cambridge 1977.

- [53] T. Sejnowski, P. Churchland, *Brain and Cognition*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [54] H. Simon, C. Kaplan, *Foundations of Cognitive Science*, w: *Foundations of Cognitive Science*, red. M. Posner, Bradford Books/MIT Press, Cambridge, MA 1989.
- [55] B. Skinner, *Beyond Freedom and Dignity*, Alfred A. Knopf, Nowy Jork 1971.
- [56] P. Thagard, *Computational Philosophy of Science*, Bradford Books/MIT Press, Cambridge, MA 1986.
- [57] N. Thomas, *Coding Dualism: Conscious Thought Without Cartesianism*, Internet 1991.
- [58] T. Tomaszewski (red.), *Psychologia ogólna*, t. IV, Wydawnictwo naukowe PWN, Warszawa 1995.
- [59] J. Townsend, J. Busemeyer, *Dynamic Representation of Decision-Making*, w: *Mind as Motion*, red. R. Port, T. van Gelder, Bradford Books/MIT Press, Cambridge, MA 1995.
- [60] S. Triger, Solipsism, *Individualism and Cognitive Science*, w: "Journal of Experimental and Theoretical Artificial Intelligence", Nr 3/1991, ss. 163-170.
- [61] A. Turing, *Computing Machinery and Intelligence*, w: *Computers and Thought*, red. E. Feigenbaum, J. Feldman, MIT Press, Cambridge, MA 1995.
- [62] J. Watson, *Behawioryzm. Psychologia jak ją widzi behawiorysta*, PWN, Warszawa 1913, 1990.
- [63] A. Wierzbicka, *Emotion and facial expression: A semantic perspective*, w: "Culture and Psychology", Nr 1/1995, ss. 227-258.
- [64] R. Wilson, Individualism, w: *The MIT Encyclopedia of Cognitive Science*, red. R. Wilson, F. Keil, MIT Press, Cambridge, MA 1999.
- [65] T. Winograd, *Formalisms for Knowledge*, w: *Thinking: Readings in Cognitive Science*, red. P. Johnson, P. Wason, Cambridge University Press, Cambridge 1974.
- [66] A. Wróbel, *Neuron i sieci neuronowe*, w: *Mózg a zachowanie*, red. T. Górski i inni, Wydawnictwo Naukowe PWN, Warszawa 1997 (a).
- [67] A. Wróbel, *W poszukiwaniu integracyjnych mechanizmów działania mózgu*, w: *Mózg a zachowanie*, red. T. Górski, i inni, Wydawnictwo Naukowe PWN, Warszawa 1997 (b).