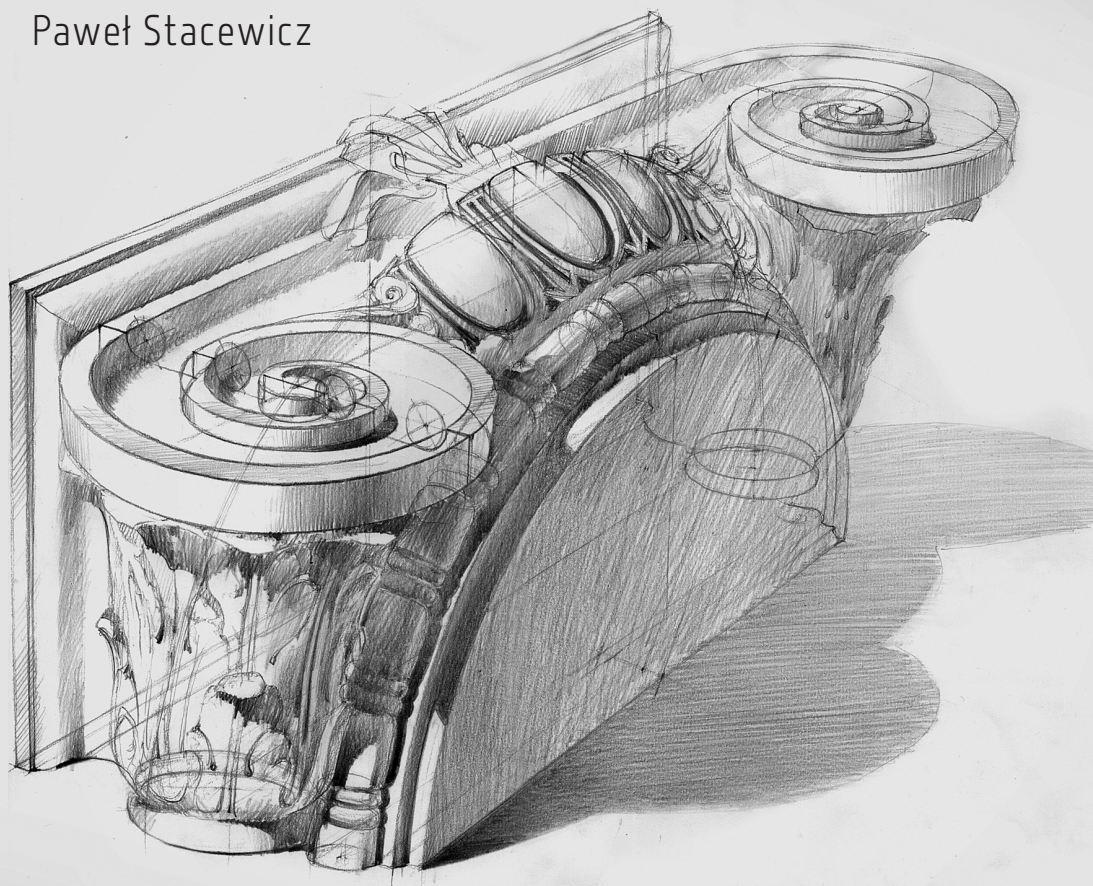


INFORMATYKA A FILOZOFIA

Pojęcia jako funkcje decyzyjne

Zagadnienia filozoficzne, metodologiczne i informatyczne

Paweł Stacewicz



CALCU LEMUS

OFICyna WYDAWNICZA POLITECHNIKI WARSZAWSKIEJ

INFORMATYKA A FILOZOFIA

**Pojęcia jako funkcje decyzyjne.
Zagadnienia filozoficzne,
metodologiczne i informatyczne**

INFORMATYKA A FILOZOFIA

Pojęcia jako funkcje decyzyjne

Zagadnienia filozoficzne, metodologiczne
i informatyczne

Paweł Stacewicz



CALCU LEMUS

OFICyna WYDAWNICZA POLITECHNIKI WARSZAWSKIEJ
WARSZAWA 2021

Pojęcia jako funkcje decyzyjne. Zagadnienia filozoficzne, metodologiczne i informatyczne
Wydanie I

Słowa kluczowe: pojęcia, funkcje decyzyjne, reprezentacje pojęciowe, pojęcia a zbiory, sztuczna inteligencja, nieprzejrzystość poznawcza, problem czarnej skrzynki

Recenzenci

prof. dr hab. André Włodarczyk

prof. dr hab. Mieczysław Omyła

Publikacja powstała w ramach projektu Narodowego Centrum Nauki pt. "Turing, Ashby i aktywność mózgu" (nr projektu: 2020/37/B/HS1/01809; konkurs: OPUS 19; PI: Hajo Greif, CI: Paweł Stacewicz)

Korekta

Teresa Woźniak

Projekt okładki

Danuta Czuდეk-Puchalska

Skład komputerowy wykonała Agencja Reklamowa *Efectoro*

© Copyright by Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2021

Wydawca: Politechnika Warszawska

Oficyna Wydawnicza Politechniki Warszawskiej (UIW 48800)

ul. Polna 50, 00-644 Warszawa, tel. 22 234-70-83

Księgarnia internetowa Oficyny Wydawniczej PW www.wydawnictwopw.pl

tel. 22 234-75-03; e-mail: oficyna@pw.edu.pl

Utwór w całości ani we fragmentach nie może być powielany ani rozpowszechniany za pomocą urządzeń elektronicznych, mechanicznych, kopiujących, nagrywających i innych, w tym nie może być umieszczany ani rozpowszechniany w Internecie bez pisemnej zgody posiadacza praw autorskich

ISBN 978-83-8156-274-4 (druk)

ISBN 978-83-8156-275-1 (online)

Zamówienie nr 132/2021

Druk i oprawa: Drukarnia Oficyny Wydawniczej Politechniki Warszawskiej, tel. 22 234-70-30

Spis treści

Wprowadzenie	7
Rozdział 1. Pojęcie jako podstawowy składnik ludzkiej wiedzy	10
1.1. Wiedza w sensie psychologicznym i logicznym	10
1.2. Pojęcia jako rdzeń wiedzy	13
1.3. Pojęcia i nazwy	14
1.4. Pojęcia w perspektywie psychologicznej	18
1.5. Podsumowanie	21
Rozdział 2. Pojęcia jako funkcje	23
2.1. Pojęcia-funkcje w historycznym ujęciu Gottloba Fregego	24
2.2. Interpretacja operacjonistyczna koncepcji Fregego	26
2.3. Możliwe rozszerzenia koncepcji Fregego	28
2.4. Podsumowanie	30
Rozdział 3. Teorie pojęć a teorie zbiorów	31
3.1. Zakresy i formuły pojęć-funkcji	31
3.2. Pojęcia ostre i rozmyte	33
3.3. Pojęcia pewne i przybliżone	38
3.4. Pojęcia zbiorowe i niezbiiorowe	44
3.5. Podsumowanie	47
Rozdział 4. Pojęcia w perspektywie informatycznej	48
4.1. Obszary badawcze i zagadnienia	48
4.2. Informatyczne reprezentacje pojęciowe	52
4.3. Od reprezentacji do algorytmów	58
4.4. Podsumowanie	59
Rozdział 5. Zagadnienie przejrzystości poznawczej reprezentacji pojęciowych	60
5.1. Metafora i problem czarnej skrzynki	60
5.2. Przejrzystość poznawcza reprezentacji pojęciowych	66
5.3. Różne typy i poziomy wyjaśnień	69
5.4. Podsumowanie	72
Bibliografia	73

Wprowadzenie

Czym są pojęcia, że możemy je pojmować na różne sposoby, wnikając w różne warstwy wytwarzanej i gromadzonej przez człowieka wiedzy? Pytanie to nurtuje filozofów, psychologów, semiotyków, logików, lingwistów, ale także informatyków i specjalistów od sztucznej inteligencji. Nurtuje też autora tej książki.

Tytuł książki, przez zawartą w nim frazę „funkcje decyzyjne”, wskazuje główny wątek rozważań, zogniskowany wokół zdolności systemów operujących pojęciami do podejmowania decyzji kategoryzacyjnych. Jest to w istocie najważniejsza i najbardziej podstawowa funkcja pojęć. Czy to oto urządzenie jest komputerem cyfrowym? Czy ten oto tekst należy do kategorii wprowadzających? Czy ten oto owoc jest jabłkiem? Odpowiadając na tego typu pytania, nawet jeśli nie zadajemy ich wprost, musimy odwoływać się do różnych pojęć. Operując nimi, nie tylko nazywamy i porządkujemy świat, lecz poszerzamy i pogłębiaamy zakres tego, co wiemy. Bez pojęć proces ten byłby niewykonalny: nie moglibyśmy uzyskać i utrwalić żadnej bardziej zaawansowanej wiedzy. I choć uwagi te dotyczą ludzi, w podobny sposób, na podstawie różnie reprezentowanych struktur pojęciowych, działają systemy informatyczne oparte na wiedzy.

Myślenie o pojęciach w kategoriach abstrakcyjnych funkcji – funkcji porządkujących obiektom z określonej dziedziny pewne decyzje kategoryzacyjne – zapoczątkował niemiecki logik i filozof Gottlob Frege. W wielu miejscach książki, w szczególności w rozdziale 2, pokazuję, jak różne uszczegółowienia i rozszerzenia koncepcji Fregego przeniknęły do myśli współczesnej (w tym: do informatyki). Dotyczy to w szczególności zakresów pojęć, które tradycyjnie rozumiano jako zbiory ostre, opisywane formalnie za pomocą klasycznych aksjomatyk teorii mnogości (np. Zermela-Fraenkla). Interpretując zakresy pojęć jako zbiory nieklasyczne, np. rozmyte lub przybliżone, uzyskuje się dostęp do sfery pojęć używanych przez ludzi w codziennej komunikacji. To zaś jest istotne zarówno dla lingwistów (badających język naturalny), jak i informatyków (poszukujących efektywnych rozwiązań w zakresie komunikacji człowiek-maszyna). Kwestie te omawiam wstępnie w rozdziale 3, a rozwijam w rozdziale 4.

Drugie rozszerzenie koncepcji Fregego polega na wnikięciu w postaci formuł określających pojęcia-funkcje. Odpowiednich narzędzi do tego dostarcza informatyka, a konkretniej, proponowane na potrzeby sztucznej inteligencji metody reprezentacji wiedzy. Są to metody formalne, możliwe do implementacji w różnych fizykalnie systemach, co powoduje, że uprawiana przy ich użyciu teoria pojęć (w tym: ich możliwych reprezentacji) nadal pozostaje abstrakcyjna i z matematyzowana. Wybrane metody tego rodzaju przedstawiam w rozdziale 4.

We współczesnych zastosowaniach informatyki, w związku z własnościami niektórych reprezentacji pojęciowych, występuje pewien niepokojący problem, nazywany problemem czarnej skrzynki (ang. *black box problem*). Dotyczy on zarówno systemów zorientowanych praktycznie, na pewne konkretne zadania inżynierskie, jak i układów modelujących ludzką aktywność poznawczą. Problem polega na tym, że systemy skutecznie operujące pojęciami (to znaczy: dochodzące do poprawnych lub pożądaných wyników) nie potrafią przedstawić człowiekowi przejrzystych znaczeniowo zależności, które wyjaśniają ich decyzje.

Głównym, aczkolwiek nie jedynym, źródłem problemu czarnej skrzynki jest konieczność wbudowywania w systemy sztuczne różnych mechanizmów uczenia się, czyli automatycznego generowania wiedzy na podstawie dostarczanych przez użytkownika przykładów lub samodzielnych doświadczeń. Owa konieczność może wynikać z potrzeb czysto praktycznych i/lub z intencji modelowania w systemie zdolności typowo ludzkich (wśród których szczególne miejsce zajmuje zdolność do adaptacji i uczenia się). Niektóre z informatycznych metod reprezentacji wiedzy, głównie metody subsymboliczne, używane w sztucznych sieciach neuronowych, wymuszają taki sposób uczenia się, który nie daje wglądu w wysokopoziomowe, dostępne dla zwykłego użytkownika, uzasadnienia podejmowanych decyzji. Ma to miejsce na przykład wtedy, gdy uczenie się polega na dostosowywaniu pewnych technicznych, generowanych roboczo parametrów do wymuszanych na systemie decyzji końcowych. System „nauczony i działający” przypomina wówczas nieprzejrzystą znaczeniowo czarną skrzynkę o bardzo ograniczonej zdolności generowania wyjaśnień. W odniesieniu do takich systemów powstają doniosłe filozoficzne pytania o status dostarczanej przez nie wiedzy.

Mam nadzieję, że naszkicowane powyżej główne wątki książki zachęciły potencjalnego czytelnika do zagłębienia się w poszczególne fragmenty lub złożoną z nich całość. Zgodnie z nazwą serii wydawniczej, *Informatyka a filozofia*, tematyka i forma książki są owocem połączenia dwóch badawczych perspektyw: informatycznej, zawężonej dodatkowo do tematyki sztucznej inteligencji, oraz filozoficznej, zogniskowanej wokół zagadnień epistemologicznych. Najważniejsza myśl książki jest następująca: konstytuujące naszą wiedzę pojęcia warto ujmować abstrakcyjnie, formalnie i algorytmicznie, bo

to daje nam cenny filozoficznie wgląd w ich naturę, strukturę i dostępność poznawczą¹.

Na koniec chciałbym serdecznie podziękować dwóm recenzentom monografii, Profesorowi Mieczysławowi Omyle oraz Profesorowi André Włodarczykowi, za cenne dyskusje i uwagi, które przyczyniły się istotnie do udoskonalenia wielu omówień, argumentów i sformułowań. Dziękuję też uczestnikom Seminarium z filozofii nauki przy Wydziale Administracji i Nauk Społecznych Politechniki Warszawskiej, z którymi miałem okazję dyskutować, podczas seminaryjnych posiedzeń, niektóre fragmenty książki.

¹ W tym miejscu warto wyjaśnić, że większa część książki, z wyłączeniem rozdziału 5, jest oparta na niepublikowanych dotychczas fragmentach mojej rozprawy doktorskiej, pt. „Umysł a modele maszyn uczących się. Wybrane konsekwencje filozoficzne badań nad sztuczną inteligencją”. Inne fragmenty tej rozprawy zostały opublikowane już wcześniej w monografii (Stacewicz, 2010), poświęconej modelowaniu oraz sztucznej realizacji wybranych funkcji ludzkiego umysłu za pomocą informatycznych systemów uczących się.

Rozdział 5 monografii oraz niektóre fragmenty pozostałych rozdziałów zostały opracowane w ramach realizacji projektu badawczego OPUS 2019 Narodowego Centrum Nauki (NCN), pt. „Turing, Ashby i aktywność mózgu” (nr projektu 2020/37/B/HS1/01809; PI Hajo Greif, CI Paweł Stacewicz).

Rozdział 1

Pojęcie jako podstawowy składnik ludzkiej wiedzy

Precyzyjne znaczenia terminu „pojęcie” określa się w językach nauk szczegółowych, takich jak psychologia, logika czy informatyka. Mimo to już w języku potocznym występuje szereg typowych wyrażeń i zwrotów, które są stosunkowo bliskie jego znaczeniom specjalnym. Na przykład, używa się powszechnie sformułowania „nie mieć o czymś pojęcia”. Ma ono wydźwięk ogólny, a wyraża brak zrozumienia, brak wiedzy lub brak wyrobionego poglądu na coś. Mówi się także – nieco staroświecko, lecz tym samym dobitnie – „Nie pojmuję”. To zaś znaczy tyle, co „Nie rozumiem”. Często też, przeprowadzając pewnego typu rozumowania, używa się zwrotów typu „Zgodnie z pojęciem X -a...” albo „Z samego pojęcia X -a wynika, że...”. Przykładowo, gdy ktoś twierdzi, że „Z samego pojęcia liczby podzielnej przez 6 wynika, że jest ona podzielna przez 2 i przez 3”, to przedstawia skrótowo rozumowanie polegające na wyprowadzeniu z aksjomatów teorii liczb oraz definicji dotyczących podzielności liczb konkluzji zawartej w drugiej części zdania.

Przytoczone przykłady, jak również inne intuicje językowe, pozwalają wstępnie zwrócić uwagę na ważną cechę terminu „pojęcie”: używa się go zwykle, gdy mówi się o procesach rozumienia, te zaś polegają na uchwytowaniu istoty pewnego obiektu, np. na określaniu jego istotnych cech. Spostrzeżenie to znajduje swoją egzemplifikację w dziedzinie pojęć nominalnych, które definiuje się jako znaczenia, czyli sposoby rozumienia nazw generalnych. Nazwy takie stanowią podstawowy element większości wypowiedzi, również tych, które wyrażają przekonania lub wiedzę podmiotu poznawczego.

1.1. Wiedza w sensie psychologicznym i logicznym

Pojęcie wiedzy należy rozpatrywać, licząc się z trafnością sformułowań typu „ X wie, że Z ”, a zatem uwzględniając zarówno rolę podmiotu „coś-wiedzącego”, jak

i treść zdania Z, które opisuje jego wiedzę². Mówiąc o roli podmiotu, mamy na myśli przede wszystkim jego przeżycia związane z posiadaniem wiedzy – np. poczucie pewności czy oczywistości żywionych przekonań. Przyjmujemy też, że stosunek do nich wyznacza różne pojęcia wiedzy. I tak, marginalizując rolę przeżyć myślowych podmiotu, a wysuwając na pierwszy plan normy poprawności procedur wiedzytwórczych oraz logiczny układ ich rezultatów, uzyskuje się pojęcie *wiedzy w sensie logicznym*. Uwydatniając rolę przeżyć myślowych podmiotu, a w szczególności czyniąc je gwarantem pewności, uzyskuje się pojęcie *wiedzy w sensie psychologicznym*.

Do oznaczenia podstawowego składnika wiedzy w sensie psychologicznym proponujemy zarezerwować termin „przekonanie” (i postawić go w opozycji do terminu „stwierdzenie”). W przeciwieństwie do stwierdzeń, formułowanych zawsze w pewnym języku publicznym, przekonania są zawsze „umocowane” w pamięci (szerzej: psychice) konkretnego podmiotu. Jako takie zaś **nie muszą być ani precyzyjne, ani w pełni przekładalne na język ogółu, ani powiązane ze sobą siłą merytorycznych uzasadnień**. Za ich uformowanie odpowiadają różnorodne czynniki psychologiczne, takie jak indywidualne upodobania, łatwość lub trudność przyswajania pewnych przekonań czy przebieg uczenia się³. Ważnym składnikiem przekonań są przedstawienia (naoczne i nienaoczne), których rola polega bądź na bezpośrednim wyrażaniu określonych treści, bądź na ich wzmacnianiu swoją sugestywnością. Na przykład, obrazowe interpretacje twierdzeń matematycznych mogą – zwłaszcza wobec nieznajomości odpowiednich dowodów – podtrzymywać czyjaś wiarę w ich prawdziwość. Główną funkcją przekonań pozostaje utrzymanie pożądanych relacji jednostki z otoczeniem (a także z samą sobą); jeśli zatem psychika jednostki nie jest ukierunkowana na dążenie do wiedzy obiektywnej, zawarte w niej przekonania nie muszą gwarantować skuteczności takiego dążenia. Wszystko to powoduje, że przekonania nie należy traktować jako po prostu fałszywych lub prawdziwych, ale raczej jako w pewnym stopniu wiarygodnych (dla pewnej osoby). Co ważne też, ów stopień wiarygodności zależy często od względów pozamerytorycznych, a zatem nie daje się utożsamić z żadną obiektywną miarą uzasadnienia twierdzeń na podstawie mniej lub bardziej pełnych i pewnych danych empirycznych (a także: przyjętych na poziomie teorii założeń, definicji czy aksjomatów). Por. (Chisholm 1994), (Marciszewski 1972).

Przejdźmy z kolei do charakterystyki wiedzy w sensie logicznym. Na wiedzę taką składają się stwierdzenia (zdania) nie tylko uznawane przez kogoś, ale również sformułowane w pewnym ścisłym języku i wzajemnie się *uzasadniające* [Jadacki 1985, s. 45]. Podstawę uzasadniania stanowią, z jednej strony, postulaty znaczeniowe, czyli definicje i aksjomaty przyjmowane ze względu na pewne

² Z tego względu pojęcie wiedzy (podmiotowe) należy odróżnić od pojęcia prawdy (apodmiotowego).

³ Pojęcie *przekonania* wchodzi w skład klasycznej definicji wiedzy jako prawdziwego i uzasadnionego przekonania. Por. (Chisholm 1994, s. 173-174).

kryteria (np. oczywistość), a z drugiej strony, zdania empiryczne, przyjmowane z uwagi na zgodność ich treści z opisywaną rzeczywistością. Poprawności uzasadniania strzegą normy formułowane w logice ogólnej, zwłaszcza w metodologii nauk i epistemologii (Kotarbiński 1986). **To właśnie fakt przestrzegania tychże norm, norm logicznych, pozwala nazwać fragment żywionych przez kogoś przekonań wiedzą w sensie logicznym.** Jeśli chodzi o precyzję i obiektywność, jej ideał stanowi *wiedza sformalizowana*, czyli zbiór zinterpretowanych formuł pewnej teorii formalnej, np. rachunku predykatów pierwszego rzędu⁴. Oczywiście wiedza ta – ze względu na czysto psychologiczne uwarunkowania jej potencjalnych użytkowników – nie musi być dla wszystkich tak samo dostępna, np. równie czytelna lub równie łatwo przyswajalna; nie zawsze też musi sprawdzać się w procesie międzyludzkiej komunikacji⁵.

Zarysowanej dwuznaczności pojęcia wiedzy odpowiada dwojaka interpretacja jej podstawowego składnika, tj. pojęć. Pojęcia bowiem rozumie się **bądź psychologicznie, bądź logicznie**, a rozróżnienie to zyskało sobie popularność zwłaszcza wśród przedstawicieli i kontynuatorów Szkoły Lwowsko-Warszawskiej. Za Kazimierzem Ajdukiewiczem możemy przytoczyć takie oto objaśnienie tej kwestii: *„Myśli wypowiedane za pomocą nazw należą do tego rodzaju przeżyć, które nazywa się w psychologii przedstawieniami. Psychologia rozróżnia wśród przedstawień tzw. przedstawienia naoczne, czyli wyobrażenia, i przedstawienia nienaoczne, które w psychologii nazywa się pojęciami. (...) Terminu „pojęcie” używa się jednak także w innym sensie, który nazywać będziemy jego sensem logicznym. Mianowicie pojęciem w sensie logicznym, a w szczególności pojęciem nominalnym w sensie logicznym, nazywamy znaczenie jakiejś nazwy”* (Ajdukiewicz 1965, s. 30).

Objaśnienie to wiąże pojęcia ze sferą języka, który służy do ich zapisywania i komunikowania. W podrozdziale 1.3 wkroczymy w tę sferę głębiej, przyglądając się bliżej językowemu nośnikowi wspomnianych powyżej pojęć nominalnych, czyli nazwom.

⁴ Od wiedzy sformalizowanej należy odróżnić wiedzę formalną: o ile ta pierwsza ma charakter zinterpretowany, tj. stosuje się do konkretnych obiektów, o tyle ta druga polega na znajomości samych (niezinterpretowanych) formuł teorii formalnej i reguł posługiwania się nimi. Z tego względu obejmuje, na przykład, zagadnienia składni danej teorii oraz semantyczne teorie jego potencjalnych interpretacji.

⁵ Zaznaczmy tu, że obecnie granica między opisem wiedzy w sensie logicznym (ze swojej natury: opisem formalnym) i psychologicznym zaczyna się zacierać. Między innymi za sprawą informatyki – stosowanej między innymi do modelowania ludzkich czynności poznawczych – język opisujący strukturę ludzkiej pamięci i zawartych w niej przekonań staje się równie precyzyjny jak symbolika wiedzy sformalizowanej. Jest to bowiem albo język ogólnie pojętej algorytmiki, albo jakiś konkretny język programowania (czyli kodowania algorytmów na potrzeby ich konkretnej, komputerowej implementacji). W ten sposób psychologia (w tym przypadku: poznawcza) zbliża się do ujęć metodologiczno-formalnych.

1.2. Pojęcia jako rdzeń wiedzy

Wskazanie pojęć jako podstawowego składnika wiedzy – zarówno w sensie logicznym, jak i psychologicznym – uznajemy za słuszne i ze względów strukturalnych, i ze względów funkcjonalnych.

Rozważmy na początek strukturę wiedzy. W swojej warstwie językowej stanowi ona pewien mniej lub bardziej sformalizowany tekst, który składa się ze zdań. Te zaś nabierają sensu, czyli stają się *sądami* w sensie logicznym, dopiero wtedy, gdy zawierają jakieś nazwy lub dotyczą, chociażby kontekstowo, pewnych obiektów (identyfikują je lub opisują). W szczególności, gdy skupimy uwagę na wiedzy sformalizowanej w sposób uwzględniający strukturę zdań atomowych, to musimy spostrzec, że w każdym zdaniu atomowym o ogólnej postaci *A jest B* (tzw. zdania podmiotowo-orzecznikowe) występują dwie nazwy, *A* i *B*. Pamiętając więc, że mówimy o pojęciach nominalnych będących znaczeniami nazw, możemy stwierdzić, co następuje: **tak jak nazwy stanowią konieczny element zdań, tak pojęcia stanowią konieczny element sądów.**

Do podobnej konkluzji skłania obserwacja praktyki badawczej psychologów. Wszak znakomita większość proponowanych przez nich modeli pamięci semantycznej, nazywanej niekiedy „wiedzą, że”⁶, zawiera jako swój trzon hierarchię lub sieć pojęć. Typowy model, wykorzystywany także w informatyce, nosi nazwę „sieci semantycznej”. Składa się on, po pierwsze, z węzłów reprezentujących cechy, pojęcia i desygnaty pojęć, po drugie, z połączeń między węzłami, które reprezentują relację między wskazanymi obiektami (takie jak relacja typu pojęcie-cecha, np. „stół jest okrągły”; czy relacja typu pojęcie-pojęcie, np. „stół jest jednym z mebli”). Por. (Lindsay, Norman 1984).

W aspekcie funkcjonalnym wiedza stanowi podstawę różnych czynności poznawczych, takich jak przypominanie sobie, wnioskowanie czy uczenie się. Z wiedzą pojęciową najbardziej bezpośrednio wiąże się *percepcja*, polegająca między innymi na rozpoznawaniu i klasyfikowaniu obiektów. Pojęcia – rozumiane przez psychologów jako uogólnione schematy pamięciowe⁷ grup obiektów podobnych – pozwalają, po pierwsze; **identyfikować obiekty jako przynależne do określonych kategorii**, po drugie, **potwierdzać tożsamość pojedynczych obiektów, które zmieniają się w czasie lub są postrzegane w różnych okolicznościach**. Dzięki pojęciom podmiot percepcji nie musi w sposób nadmierny angażować swoich zasobów pamięciowych:

⁶ Wiedza ta ma charakter ogólny, dotyczy świata zewnętrznego wobec jednostki, jest uporządkowana relacyjnie (a nie czasowo), najczęściej ma wyraz językowy i zakłada możliwość zweryfikowania pod kątem prawdziwości. (Tulving 1983; Maruszewski 1986).

⁷ Konkretny schemat, obecny w pamięci konkretnej osoby, nazywa się w psychologii reprezentacją pojęciową (czyms co konkretnie reprezentuje abstrakcyjne pojęcie). W psychologii poznawczej rozważa się różne typy reprezentacji, opisując je na różnych poziomach (np. semantycznym – dotyczącym pamięci, czy neuronalnym – dotyczącym mózgu i systemu nerwowego). Więcej informacji na ten temat zawiera podrozdział 1.4.

zamiast przechowywać w pamięci bardzo wiele podobnych reprezentacji poszczególnych obiektów, tworzy i wykorzystuje jedną reprezentację zbiorczą. Na przykład, chcąc poprawnie zidentyfikować czytane właśnie litery polskiego alfabetu, nie trzeba przechowywać w pamięci ich niezliczonych kształtów w różnych czcionkach, wystarczy utrwalić i wykorzystywać jeden uogólniony wzorec. Z czynnościami percepcyjnymi wiąże się inna ważna i poznawczo elementarna funkcja pojęć, a mianowicie **określanie istotnych własności obiektów**. Dzięki temu, że ludzie dysponują nie izolowanymi pojęciami, ale pewną ich uporządkowaną hierarchią lub siecią, przeprowadzają tę operację automatycznie i bez wysiłku (Maruszewski 1996). Na przykład, aby stwierdzić, czy ten oto samochód marki Ford ma pewną własność (np. ma silnik lub koła), wystarczy odwołać się do własności klasy, do której należy (czyli klasy samochodów). Klasę tę reprezentuje osobne pojęcie.

Określanie własności obiektów przez odwołanie się do układu pojęć można uznać za najprostszą formę rozumowania. Co więcej, forma ta jest niezbędna do tego, aby przeprowadzać kolejne, bardziej zaawansowane, operacje myślowe – na przykład: na podstawie dwóch sądów o własnościach pewnych obiektów wyprowadzać sąd trzeci (co umożliwiałaby sylogistyka). Ze wskazanych względów same pojęcia odgrywają podstawową rolę nie tylko w percepcji, ale również w *rozumowaniu*. *Notabene*, o tym samym przekonuje dokonana wcześniej analiza struktury wiedzy: wszak wszelkie sądy, będące punktem wyjścia lub wynikami rozumowań, zawsze zawierają jakieś pojęcia.

1.3. Pojęcia i nazwy

Punktem wyjścia semiotyczno-logicznych rozważań nad pojęciami czyni się zwykle ogólnikowe sformułowanie, iż „*pojęcie to znaczenie nazwy*” (Ajdukiewicz 1965; Kotarbiński 1986). Kryje się za nim słuszne założenie, iż znaczenie konkretnej nazwy wykracza poza jej kształt i brzmienie, a stanowi raczej coś, co przysługuje wspólnie nazwom o różnych kształtach i brzmieniach; coś, co sprawia, iż odnoszą się one w ten sam sposób do tych samych obiektów (ewentualnie: są w taki sam sposób używane). Z tego względu należy odróżnić nazwę (taki a taki obiekt fizyczny) od jej znaczenia, czyli pojęcia. Tutaj jednak wynika problem uszczegółowienia prawej strony ogólnikowej tożsamości „*pojęcie = znaczenie nazwy*”. Problem ten wyraża się w istocie dwoma pytaniami: (1) „Czy i jak znaczenie zależy od rodzaju danej nazwy?”, oraz (2) „Jakie własności semiotyczne nazw składają się na ich znaczenie?”.

Odnosząc się do pytania drugiego, trzeba stwierdzić, że na polu semiotyki wypracowano bardzo wiele, istotnie różnych koncepcji znaczenia. Przekonującą próbę ich uporządkowania znajdujemy w pracy Jerzego Pelca, „O użyciu wyrażań” (Pelc 1971). Zaznaczając, iż proponowany podział nie jest ani rozłączny, ani wyczerpujący, Pelc **dzieli teorie znaczenia na referencyjne i operacyjne**. Na gruncie

pierwszych przyjmuje się wstępnie, że właściwa odpowiedź na pytanie „Co znaczy nazwa N?” musi wyjaśniać, do czego i jak nazwa N się odnosi; w najprostszym wypadku zaś – co oznacza. Zacytujmy: „...znaczenie wyrażenia upatruje się albo w pozajęzykowym odpowiedniku tego wyrażenia (...), albo w relacji między wyrażeniem a jego odpowiednikiem pozajęzykowym.” (Pelc 1971, s. 204). Do tak scharakteryzowanej grupy zaliczają się między innymi teorie J. Locke’a (znaczeniem słów są pewne idee, czyli pozajęzykowe przedmioty idealne), J.S. Milla i Fregego (znaczenie nazwy stanowi jej konotacja, czyli zespół cech charakterystycznych dla wszelkich jej desygnatów), Husserla (znaczeniem słów są przedmioty intencjonalne) i Ch. Morrisa (znaczenie nazwy stanowi typ reakcji, które za jej pomocą mogą być potencjalnie wywołane u słuchaczy). W przypadku teorii drugiego rodzaju, teorii operacyjnych, podkreśla się, że pytanie o znaczenie nazwy jest w istocie pytaniem o dozwolone jej zastosowania (dozwolone ze względu na reguły danego języka). Znaczenie konkretnej nazwy polega więc na tym, że może ona występować w takich a takich konstrukcjach składniowych (a w innych nie) i nadaje się do użycia w takich a takich sytuacjach (a w innych nie). Co najważniejsze jednak, użycie nazwy w konkretnej sytuacji jest w stanie spowodować określone zmiany, np. w przeżyciach lub zachowaniu odbiorcy wypowiedzi, albo we fragmencie świata, którego dotyczy wypowiedź. Autor referowanej klasyfikacji proponuje taką oto charakterystykę: „(PS: W teoriach operacyjnych) utożsamia się znaczenie z różnie rozumianymi: użyciem (use) lub sposobem użycia (usage) wyrażenia, zwłaszcza zaś zdania rozważanego w kontekście językowym i na tle sytuacji pozajęzykowej, traktując język jako narzędzie, a mowę jako pewnego rodzaju działanie wykonywane za pomocą narzędzi językowych czy jako grę językową”. (Pelc 1971, s. 204). Za pierwszego zdecydowanego zwolennika podejścia operacyjnego uznać trzeba L. Wittgensteina (jako autora *Dociekań filozoficznych*); do innych wpływowych jego propagatorów zalicza się J.L. Austina, G. Ryle’a i P.F. Strawsona.

Dodatkowym źródłem komplikacji przy tworzeniu teorii znaczenia jest kwestia **różnych typologii nazw**, a dokładniej: niezgoda semiotyków co do tego, jakie kategorie nazw należy odróżnić i tym samym, w jaki sposób należy uszczegółowić ogólną charakterystykę pojęć (mówiąc o uszczegółowieniu, mamy na myśli przejście od pojęć jako znaczeń nazw w ogóle do pojęć jako znaczeń nazw takich a takich). Ponieważ ani niniejszy ustęp, ani cała praca nie mają charakteru specjalistyczno-semiotycznego, w zagadnienie to nie będziemy wnikać⁸. Przypomnimy tylko, że jego rozwiązanie podręcznikowe, a przy okazji rozpowszechnione wśród niespecjalistów,

⁸ Zaznaczmy jednak, że zagadnienie adekwatnej typologii nazw można rozumieć bądź sprawozdawczo, bądź normatywnie. W ujęciu sprawozdawczym chodzi o to, by ustalić, jakie klasyfikacje nazw są najbardziej płodne ze względu na zdanie sprawy z funkcji nazw w języku naturalnym (tj. w rzeczywistej praktyce językowej pewnej społeczności). W ujęciu normatywnym chodzi natomiast o co innego, a mianowicie o zaproponowanie takiej typologii, które gwarantowałyby owocne badania nad pewnymi językami sztucznymi – językami bądź wyabstrahowanymi z języka naturalnego w celu uściślenia przeprowadzanych w nim rozumowań (językami logiki), bądź językami utworzonymi na potrzeby danej nauki (np. fizyki).

wygląda następująco. Ze względu na liczbę desygnatów rozróżnia się nazwy *puste*, *jednostkowe* i *ogólne*; ze względu na sposób odnoszenia się do desygnatów – *generalne* (konotujące) i *indywidualne* (niekonotujące); ze względu na strukturę wewnętrzną desygnatów – *zbiorowe* i *niezbiorowe*; ze względu na jednoznaczność wskazywania desygnatów – *ostre* i *nieostre*. Por. (Ziemiński 2000).

Wróćmy jednak do ogólnej kwestii znaczenia. Spośród dwóch wspomnianych wcześniej typów teorii – referencyjnych i operacyjnych – autorowi niniejszej pracy bliższe jest referencyjne rozumienie znaczenia nazw, a dokładniej rozumienie go jako pewnego **kompleksu własności semiotycznych**, które są ponadto uporządkowane ze względu na swoją doniosłość; dwa pierwsze miejsca takiego kompleksu zajmują kolejno: własność oznaczania i konotowania. Z takiego punktu widzenia za równoznaczne uznaje się przede wszystkim nazwy równozakresowe, ale ponadto – ponieważ samo oznaczanie nie zawsze wystarcza – mające szereg innych wspólnych własności semiotycznych, takich jak tożsamość treści i tożsamość kontekstu językowego, w którym nazwa występuje (np. czy wchodzi ona w skład zdania oznajmującego, pytającego, czy rozkazującego). Uważamy też, że znaczenie samych nazw – w odróżnieniu od znaczenia zwrotów zawierających nazwy – ujawnia się najpełniej w zdaniach typu „ x jest N ”, gdzie x reprezentuje bądź wprost obiekt wskazywany, bądź nazwę indywidualną o znaczeniu określonym ostensywnie (przez wskazanie). Jeśli pewna nazwa N występuje w zdaniu innego rodzaju, zdanie to należy przereklamować do takiej postaci, iżby znalazły się w nim, oprócz kwantyfikatorów i spójników, wyłącznie zdania typu „ x jest N ”. Na przykład zdanie „Płóć jest rybą” należy przekształcić do postaci „Dla wszelkiego x : jeśli x jest płocią, to x jest rybą”, a zdanie „Jan łapie ryby nad rzeką” do postaci „Jan jest osobą, która łapie ryby nad rzeką”⁹.

⁹ Zdania podmiotowo-orzecznikowe typu „ A jest B ” stanowią dobry punkt wyjścia do poszukiwań adekwatnej typologii nazw w ujęciu normatywnym. Decyzje co do wyboru takiej a nie innej typologii jako naczelną zależą od interpretacji łącznika *jest*.

Gdy przyjmie się założenie, iż słówko *jest* należy rozumieć jako „*jest jakieś*”, wtedy za A można wstawiać tylko nazwy bez konotacji, czyli imiona własne, za B natomiast tylko nazwy konotujące, czyli generalne. Automatycznie zatem pojawia się rozróżnienie nazw na generalne (nadające się wyłącznie na orzeczniki) oraz indywidualne (nadające się wyłącznie na podmioty). Wtedy też znaczenie przypisuje się albo tylko nazwom generalnym (dopuszczając utożsamienie znaczenia z treścią nazwy), albo obu typom nazw – zaznaczając jednak, że na czym innym polega znaczenie nazw generalnych, a na czym innym indywidualnych.

Gdy natomiast przyjmie się założenie, iż słówko *jest* należy rozumieć jako „*jest jednym z*”, wtedy zdania „ A jest B ” uznaje się za zawsze sensowne, lecz warunkiem koniecznym ich prawdziwości czyni się wymaganie, iż w miejscu A musi stać nazwa o jednym desygnacie, a w miejscu B nazwa o co najmniej jednym desygnacie. Automatycznie zatem pojawia się rozróżnienie nazw tylko ze względu na liczbę desygnatów, czyli na puste, jednostkowe i ogólne. Konsekwentnie zaś, skoro rozróżniając nazwy, nie bierze się pod uwagę ich konotacji, znaczenie nazwy utożsamia się z takim a nie innym zestawem jej desygnatów. Ewentualnie, gdyby ktoś chciał w ramach ujęcia takiego wyrażać znaczenie również za pomocą konotacji, musiałby uznać iż wszelkie nazwy mają jakąś konotację (nie ma więc wyróżnionej ze względu na znaczenie klasy nazw indywidualnych; np. nazwa „Rysy” ma taką samą konotację jak nazwa „Najwyższy szczyt w polskiej części Tatr”). (Por. (Kotarbiński 1986)).

Z punktu widzenia semiotyki logicznej najważniejszym sposobem konstytuowania znaczeń, czyli tworzenia pojęć, jest formułowanie precyzyjnych definicji. Definicje takie mogą odnosić się bądź do zastanej (a nie dość jasnej) sytuacji w języku – mają wówczas charakter *sprawozdawczy* lub *regulujący*, bądź do rzeczywistości kreowanej – mają wówczas postać *projektów*, użytecznych ze względu na pewne cele (np. stanowią fragment budowy języka pewnej nauki). Niezależnie od typu definicji jej funkcja znaczeniowótórcza polega na jednoznacznym określeniu wybranej własności semiotycznej nazwy. Z tego punktu widzenia rozróżnia się definicje *ostensywne* – polegające na bezpośrednim wskazaniu lub wyliczeniu desygnatów nazwy, a zatem precyzujące własność oznaczania, definicje *analityczne* – polegające na określeniu treści nazwy, a zatem precyzujące własność konotowania, i definicje *kontekstowe* (w uwikłaniu) – określające znaczenia szerszych zwrotów, najczęściej zdań, w których nazwa występuje (por. (Ajdukiewicz 1965)).

Konkludując zatem: choć większość semiotyków zgadza się co do tego, że należy odróżniać nazwę, mogącą przyjmować różne kształty i brzmienia, od jej znaczenia, czyli pojęcia, to nie udało im się wypracować ani jakiejś jednolitej i powszechnie przyjmowanej teorii znaczenia, czyli teorii pojęć nominalnych, ani jednolitego stanowiska wobec klasyfikacji nazw, czyli rodzajów pojęć. Niezależnie od tego w ramach ujęcia normatywnego sformułowano w semiotyce szereg dyrektyw poprawnego definiowania terminów. Ponieważ dyrektywy te, z jednej strony, pokazują, jak w idealnym wypadku powinno się określać znaczenia nazw i zawierających je zwrotów, a z drugiej strony, służą do określania różnych własności semiotycznych nazw, dostarczają one dobrego uzasadnienia tezy, iż **znaczenie nazwy stanowi kompleks jej własności semiotycznych o różnej doniosłości**.

Tezę tę rozwinął w *Logice pragmatycznej* Kazimierz Ajdukiewicz (Ajdukiewicz 1965). To, co nazwaliśmy powyżej kompleksem własności semiotycznych, Ajdukiewicz określał jako istotne względy, pod którymi rozpatrujemy i precyzujemy znaczenie nazwy, czyli przyjęty w danym języku *sposób jej rozumienia*. Pełnej listy tych względów nie wymienił, lecz wskazał przykładowo na cztery jej elementy: (1) wzgląd na przedmioty określone nazwą, (2) wzgląd na metodę rozstrzygania, czy dany przedmiot można określić nazwą, (3) wzgląd na odniesienie mówiącego do nazwy, (4) wzgląd na stosunek emocjonalny mówiącego do nazwy. Według tej koncepcji dwie osoby używające tej samej nazwy rozumieją ją w taki sam sposób (tj. dysponują takim samym pojęciem), jeśli ich myśli zgadzają się przynajmniej pod wyłuszczonej powyżej względami. A zatem jeśli: (1) chodzi im o ten sam przedmiot lub te same przedmioty, (2) posługują się tą samą metodą rozstrzygania, czy dany przedmiot należy do zakresu nazwy (metoda ta może, ale nie musi, polegać na sprawdzeniu, czy przedmiot posiada cechy wchodzące w skład treści nazwy), (3) odnoszą się do nazwy w taki sam sposób (np. informujący, pytający lub

deklarujący), (4) łączą z wypowiedzaną nazwą taki sam stosunek emocjonalny (np. lekceważenie lub uznanie)¹⁰.

W odróżnieniu od sposobu rozumienia, który obowiązuje wszystkich użytkowników języka, rozumienie nazw przez konkretne osoby zasadza się, mówiąc językiem Ajdukiewicza, na pewnych myślach. Rozważając te myśli nie pod kątem tego, jak powinny wyglądać, aby zapewnić poprawność rozumowań, ale pod kątem ich rzeczywistej struktury i funkcji, jesteśmy zmuszeni porzucić logikę i wkroczyć na teren psychologii. Możemy wówczas zadać cztery następujące pytania:

- Jak są reprezentowane w naszym umyśle zbiory obiektów o wyróżnionych własnościach?
- Czy istnieje jakaś jedna forma reprezentacji, czy też jest ich wiele – wykorzystywanych w różnych sytuacjach?
- Jakie funkcje pełnią reprezentacje w naszej aktywności poznawczej?
- Jak są tworzone reprezentacje?

Próba bliższego rozważenia tych pytań prowadzi wprost do psychologicznego ujęcia pojęć, czyli teorii reprezentacji pojęciowych.

1.4. Pojęcia w perspektywie psychologicznej

W ramach badań psychologicznych pojęcia nie są traktowane jako konstrukcje poprawne logicznie, a więc regulowane lub wprowadzane za pomocą poprawnych definicji, lecz jako pewne **faktyczne, choć opisywane na gruncie pewnej abstrakcyjnej teorii, zapisy w pamięci konkretnych osób**. Dla uwydatnienia owej różnicy na polu psychologii poznawczej nazywa się je *reprezentacjami pojęciowymi*, co zresztą możemy potraktować jako inny wariant stylistyczny używanego przez filozofów określenia przedstawienia nienaoczne (termin „przedstawienie” jest tutaj odpowiednikiem terminu „reprezentacja”). W opracowaniach z zakresu psychologii poznawczej znajdujemy następujące określenia: „(reprezentacje pojęciowe) są to schematyczne reprezentacje zbiorów obiektów, określające istotne właściwości tych obiektów” (Maruszewski 1996, s. 185), albo „W psychologii pojęcie traktuje się jako uogólnioną reprezentację poznawczą, reprezentującą uogólnioną klasę obiektów (przedmiotów, zdarzeń, cech, czynności, relacji) podobnych do siebie pod pewnym względem” (Chlewiński 1999, s. 38).

¹⁰ W związku z koncepcją Ajdukiewicza nasuwają się dwie uwagi. Po pierwsze, chociaż Ajdukiewicz nie mówi tego wprost, należy przypuszczać, iż „przyjęty w pewnym języku” sposób rozumienia nazwy reguluje odpowiednia definicja. Po wtóre, warunki oznaczone powyżej jako (1) i (2) rozumie się najprawdopodobniej tak, że w danym przypadku jest spełniony albo jeden, albo drugi. Gdyby bowiem było inaczej, warunek (1) byłby zbędny – znajomość metody rozstrzygania wystarczałaby do określenia przedmiotów podpadających pod pojęcie.

Przy próbach uszczegółowienia, a następnie rozwinięcia, powyższych sformułowań wyłania się zagadnienie dwojakiego sensu terminu „reprezentacja” (Najder 1989). Z jednej strony bowiem, oznacza on pewien byt fizyczny, który daje się obserwować i w jakimś sensie mierzyć, np. strukturę neuronalną w mózgu człowieka. Z drugiej strony zaś, oznacza on pewien ścisły, a jednocześnie schematyczny opis reprezentacji fizycznej w języku pewnej nauki, na przykład logiki wzbogaconej o terminy psychologiczne. I w tym właśnie, tj. drugim, sensie należy rozumieć wyrażenie *psychologiczna reprezentacja pojęciowa*. W przypadku dość typowym przyjmuje się, że tak ujęte reprezentacje mają postać formuł logicznych wiążących w jakiś sposób cechy desygnatów pojęć. W modelach i zastosowaniach informatycznych formuły takie implementuje się jako odpowiednie struktury danych (np. drzewa czy tabele decyzyjne), co pozwala w sposób zalgorytmizowany, a więc automatyczny, operować tak czy inaczej reprezentowanymi pojęciami¹¹ (zob. rozdział 4).

Gdy skupimy uwagę na najbardziej typowym dla psychologów sposobie reprezentowania pojęć, przez odwołanie się do **istotnych własności ich desygnatów**, natknijemy się na podstawowe pytanie: „Jakie właściwości należy uznać za typowe i w jaki sposób należy się do nich odwołać?”. Odpowiedź na nie wyznacza trzy główne koncepcje reprezentacji pojęciowych. (Por. (Smith, Medin 1981; Maruszewski 1996)).

Według koncepcji najstarszej, nazywanej *klasyczną* bądź *matrycową*, reprezentacje stanowią **koniunkcje cech definicyjnych obiektów pewnej klasy**. Cechy te muszą przysługiwać wszystkim elementom tej klasy i tylko im. Sprawia to, że granica oddzielająca desygnaty pojęcia od niedesygnatów jest ostra, a zatem nie istnieją obiekty, które wchodziłyby w zakres pojęcia tylko z pewnym prawdopodobieństwem czy stopniem pewności. Do najbardziej typowych reprezentacji klasycznych należą definicje figur geometrycznych – na przykład koniunkcja cech bycia figurą płaską, zamkniętą, złożoną z punktów równo odległych od innego punktu reprezentuje wszystkie możliwe okręgi. Chociaż tradycyjnie za klasyczne reprezentacje pojęciowe uznaje się koniunkcje cech, to w pewnych rozszerzeniach tej koncepcji dopuszcza się użycie dowolnych jednoznacznych operatorów działających na wartości cech, takich jak alternatywa czy koniunkcja alternatyw. Z metodologicznego punktu widzenia, który niekoniecznie musi pokrywać się z obserwacjami psychologów, jedynym istotnym wymaganiem co do reprezentacji klasycznej jest fakt bycia definicją, a więc taką formułą, która jednoznacznie oddziela desygnaty pojęcia od niedesygnatów.

¹¹ Z punktu widzenia samej psychologii schematyzm psychologicznych reprezentacji pojęciowych rozumie się następująco. Choć zakłada się, że w indywidualnych umysłach powstają nieco inne reprezentacje konkretne, opisuje się je w sposób ogólny, tj. postuluje się pewną ich formę wspólną różnym ludziom, niezależnie od tego, czym w konkretnych przypadkach forma taka może być wypełniona. Pozostaje to zgodne z nastawieniem badawczym psychologii poznawczej (jako nauki formułującej twierdzenia ogólne).

Koncepcja klasyczna, choć oddaje dobrze specyfikę pojęć naukowych (z konieczności jednoznacznych), nie wyjaśnia w sposób wiarygodny wyników obserwacji na ludziach posługujących się zwykłym językiem. W szczególności, okazuje się, że nie wyjaśnia ona następujących efektów: efektu typowości – polegającego na tym, że ludzie wskazują pewne desygnaty pojęcia jako typowe, a inne jako nietypowe, oraz efektu cech niedefinicyjnych – polegającego na tym, że najczęściej ludzie identyfikują desygnaty pojęcia na podstawie cech niedefinicyjnych, choć wystarczająco skutecznych w praktyce. Na przykład, wiele osób spontanicznie przypisuje ssakom cechę bycia zwierzętami lądowymi i w większości przypadków trafnie wnioskuje z braku tej cechy o nieprzynależności zwierzęcia do grupy ssaków. W rzeczywistości jednak cecha ta, choć typowa, charakteryzuje nie wszystkie ssaki (wyjątek stanowią chociażby wieloryby czy delfiny).

Reprezentacje pojęciowe drugiego typu, nazywane *prototypowymi* lub *naturalnymi*, **nie spełniają wymagań ostrości**. Obiekty reprezentowane są desygnatami pojęcia w pewnym stopniu; niekiedy całkowitym, choć nie zawsze. Innymi słowy, do każdego desygnatu jest dołączony pewien stopień jego przynależności do zakresu pojęcia. Ogół desygnatów pojęcia reprezentuje jeden zbiorczy opis, tzw. *prototyp*. Najczęściej jest to lista własności najbardziej typowych desygnatów lub lista wyróżnionych własności wraz z dołączonymi do nich miarami istotności. Zaliczenie obiektu do zakresu pojęcia następuje w wyniku porównania go z prototypem – im większe podobieństwo obiektu do prototypu, tym większy stopień przynależności obiektu do zakresu pojęcia.

Reprezentacje prototypowe mają dużą wartość z punktu widzenia psychologii, ponieważ pozwalają uchwycić istotę pojęć naturalnych, stosowanych przez ludzi na co dzień, a nie tylko w okolicznościach wyjątkowych (np. w rozważaniach naukowych). Do kategorii tej należą znaczenia wielu typowych słów, stosowanych w codziennej komunikacji, takich jak „duży”, „mały”, „wysoki” czy „niski”. Choć teoria reprezentacji prototypowych wyjaśnia efekty niewytłumaczalne na gruncie koncepcji klasycznej, to ma inne mankamenty. Po pierwsze, trudności sprawia wymodelowanie relacji podrzędności-nadrzędności pojęć, co w koncepcji klasycznej przedstawia się bardzo elegancko: pojęcia podrzędne dziedziczą wszystkie cechy pojęć nadrzędnych. Po wtóre, o ile model klasyczny wspiera jednolitą i ściśle sprecyzowaną teorię zbiorów ostrych (w szczególności przewidyuje ona jednoznaczne operacje sumy i przecięcia zbiorów), o tyle podstawę koncepcji prototypowej stanowi teoria zbiorów rozmytych, która dopuszcza wiele różnych wariantów podstawowych operacji na zbiorach. Kwestię tę omówimy szerzej w rozdziale 3, w którym ukażemy związki między różnymi teoriami pojęć (reprezentacji pojęciowych) a teoriami zbiorów.

Kolejna koncepcja reprezentacji pojęciowych nosi nazwę *egzemplarzowej*. Przyjmuje się w niej, że **zakres pojęcia reprezentują jego wybrane desygnaty** (egzemplarze), a mówiąc ściślej ich opisy w kategoriach cech. Przy sprawdzaniu, czy dany obiekt podpada pod pojęcie, opisy te wykorzystuje się tak samo, jak

scharakteryzowane wcześniej prototypy. Z tego powodu zakres pojęcia jest nieostry. Reprezentacje te oddają wielkie usługi przy wstępnych charakterystykach pojęć, które dały się poznać za pośrednictwem niewielkiej liczby desygnatów. Na przykład dziecko, które zetknęło się dotychczas z jednym tylko psem Azorem, będzie przypisywać jego cechy wszystkim zwierzętom, o których ktoś opowiada, że są psami. Ponadto reprezentacje egzemplarzowe nadają się znakomicie do opisu grup obiektów, wśród których wyczuwamy jakieś ogólne podobieństwo, ale nie potrafimy sprecyzować, na czym polega. Ich wady natomiast pokrywają się z wadami reprezentacji prototypowych¹².

Zauważmy na koniec, że przedstawione powyżej koncepcje można traktować bądź jako konkurencyjne modele opisu tego samego zjawiska, bądź jako modele reprezentacji wykorzystywanych przez ludzi w różnych sytuacjach i na różnych stadiach rozwojowych. Naszym zdaniem interpretacja druga ma lepsze uzasadnienie. Prawdopodobnie też skłonność do posługiwania się określonymi reprezentacjami zależy nie tylko od natury opisywanych przez nie obiektów, ale również od predyspozycji i dojrzałości poznawczej konkretnych osób. Na przykład, umysły abstrakcyjne będą preferować reprezentacje klasyczne i prototypowe, umysły konkretne zaś egzemplarzowe. Mimo to wydaje się, że indywidualny rozwój poznawczy powinien przebiegać (i przeważnie przebiega) od używania reprezentacji egzemplarzowych, przez stosowanie nie dość jasnych, choć z grubsza zrozumiałych, reprezentacji prototypowych, aż do operowania ścisłymi reprezentacjami klasycznymi.

1.5. Podsumowanie

1. Pojęcia stanowią podstawowy i konieczny **składnik ludzkiej wiedzy**. Wiedza rozumiana psychologicznie (jako zbiór czyichś przekonań) musi obejmować pojęcia w sensie **psychologicznym**, czyli pewne nienaoczne przedstawienia lub uogólnione schematy pamięciowe zbiorów obiektów. Wiedza rozumiana logicznie – jako zbiór zdań w pewnym publicznym i sformalizowanym języku, zdań wzajemnie się uzasadniających – musi zawierać **nazwy**, których znaczenia to pojęcia (nominalne) w sensie **logicznym**.

¹² Przedstawione koncepcje są niemal symetryczne, jeśli chodzi o wady i zalety: koncepcja klasyczna wyjaśnia specyfikę pojęć naukowych i umożliwia prosty opis hierarchii pojęć (relacji podrzędności-nadrzędności), koncepcja prototypowa (a także egzemplarzowa) oddaje istotę pojęć potocznych i wyjaśnia właściwe ich stosowaniu efekty typowości/nietypowości desygnatów, nieostrości zakresów i niedefiniyjności cech charakterystycznych. Fakt ten nasunął wielu teoretykom pomysł połączenia zalet obydwu koncepcji w różnego rodzaju ujęciach syntetycznych. Por. (Najder 1993).

2. Pojęcia rozumiane jako **znaczenia nazw** (czyli pojęcia nominalne), stanowią przedmiot zainteresowania semiotyki. Wśród semiotyków wciąż trwają dyskusje co do adekwatnej **typologii nazw** i związanych z poszczególnymi typami nazw **teorii znaczenia** (teorie te można podzielić na referencyjne i operacyjne, które mają szereg dalszych wariantów).

3. Na polu psychologii poznawczej pojęcia, nazywane niekiedy **reprezentacjami pojęciowymi**, są rozpatrywane jako schematyczne opisy wyróżnionych zbiorów obiektów, które w przypadku różnych osób mogą przyjmować nieco inną postać (z perspektywy świadomych przeżyć osób używających pojęć są to po prostu ich nienaoczne, lecz uchwytnie myślowo, przedstawienia).

4. Psychologowie poznawczy rozróżniają trzy typy reprezentacji pojęciowych:
a) **klasyczne** – opisujące obiekty z pewnej klasy jednoznacznie, np. za pomocą koniunkcji ich cech definicyjnych; b) **prototypowe** – identyfikujące obiekty z pewnej klasy niejednoznacznie, np. za pomocą pewnej kombinacji cech typowych desygnatów pojęcia (tzw. prototypów); c) **egzemplarzowe** – opisujące obiekty pewnej klasy niejednoznacznie, za pośrednictwem cech wybranych desygnatów.

Rozdział 2

Pojęcia jako funkcje

Z rozważań przedstawionych w rozdziale 1 wyłania się obraz pojęć jako czegoś, co wiąże pewien fragment języka (a mianowicie zasób nazw) z obiektami pozajęzykowymi (np. konkretnymi przedmiotami). W ujęciach logiczno-semiotycznych „elementem wiążącym” jest znaczenie nazw, a w szczególności ich treść – to za sprawą określanego konwencjonalnie znaczenia wiadomo, jakie obiekty wolno określać daną nazwą. W ujęciach psychologicznych funkcję „pośredniczenia” pełnią pamięciowe, schematyczne reprezentacje zbiorów obiektów – reprezentacje takie, z jednej strony, są opatrzone etykietami językowymi (nazwami), a z drugiej strony, odnoszą do obiektów wyróżnionych pod takimi a takimi względami.

Uogólniając zatem, można powiedzieć, iż ogół pojęć – posiadanych przez jednostkę lub obowiązujących w pewnej społeczności – określa przyporządkowania nazwom ich denotacji (zbiorów desygnatów), a każde pojęcie z osobna określa przyporządkowanie konkretnym obiektom decyzji o tym, czy są one, czy też nie są desygnatami pojęcia. Ponieważ w matematyce przyporządkowania jedno-jednoznaczne nazywa się funkcjami, spostrzeżenia powyższe prowadzą do wniosku, że **najlepszym matematycznym narzędziem opisu pojęć są funkcje**¹³.

¹³ W formie przypisu warto zauważyć, że ścisła definicja funkcji – rozumianej intuicyjnie jako jednoznaczne przyporządkowanie – odwołuje się do pojęcia relacji. Funkcja jest to taka relacja dwuargumentowa $R \subseteq X \times Y$, która spełnia warunek jednoznaczności: *xRy wtedy i tylko wtedy, gdy dla każdego $x \in A$, istnieje dokładnie jeden $y \in Y$, taki że xRy* . (Ów jedyny y można zapisać jako $y(x)$, czyli „ y dla danego x ”).

Dodajmy też, że w matematyce funkcje określa się bądź (a) ogólnie – przez podanie dziedziny i zbioru wartości (np. $f: R \rightarrow (0,1)$), bądź (b) szczegółowo – poprzez (b1) podanie konkretnych par (*argument, wartość*), czyli podanie elementów relacji będącej funkcją, lub przez (b2) określenie wzoru funkcji i jej dziedziny, np. $f(n) = 1/(n^2 + 1)$, $n \in N$.

Określenia ogólne są właściwe pewnym ogólnym teoriom funkcji pewnego rodzaju (np. teorii ciągów liczbowych będących funkcjami typu $N \rightarrow R$), natomiast określenia szczegółowe pojawiają się bądź w zastosowaniach tych teorii, bądź w przykładach obrazujących ich definicje i twierdzenia.

2.1. Pojęcia-funkcje w historycznym ujęciu Gottloba Fregego

Pierwsze pomysły dotyczące ogólnego rozumienia pojęć jako funkcji zawdzięczamy Gottlobowi Fregemu – uczonemu, który położył fundamenty pod współczesne badania logiczne (Frege 1977). Punkt wyjścia rozważań Fregego można przedstawić w formie pytania: „Co to znaczy, że mamy pojęcie M-a, na przykład pierwiastka chemicznego czy pierwiastka kwadratowego czwórki?”. Znaczy to, że jesteśmy w posiadaniu metody, która określa jednoznaczne przyporządkowania różnym obiektom decyzji TAK lub NIE, a więc decyzji o tym, czy obiekt podpada pod pojęcie, czy nie podpada. Innymi słowy, nasze decyzje o przynależności obiektów do zakresu danego pojęcia są zgodne z pewną, znaną nam lub nie, funkcją¹⁴.

Zdaniem Fregego wzajemna odpowiedniość między pojęciami i funkcjami ujawnia się dobitnie na poziomie języka. Okazuje się bowiem, iż zarówno funkcje, jak i pojęcia są zapisywane za pomocą formuł niekompletnych, to znaczy takich, które wymagają uzupełnienia konkretnymi wartościami argumentów. Język tradycyjnej matematyki (np. arytmetyki) nie pozostawia co do tego żadnych wątpliwości: wszak wzory większości funkcji zawierają zmienne. Język potoczny natomiast trzeba poddać pewnej „obróbce logicznej”. Jedną z metod, dotyczącą zdań oznajmujących, polega na przekształcaniu wszelkich zdań tego typu do znanej logikom od czasów Arystotelesa postaci podmiotowo-orzecznikowej „*A jest B*”. Przykładowo: zdanie „12 dzieli się przez 3” należy zapisać jako „3 jest podzielnikiem 12”.

Jeśli zapis podmiotowo-orzecznikowy potraktuje się jako pewien uniwersalny schemat, to staje się widoczne, że przy określonym orzeczeniu *B* (np. *B* = „podzielnik 12”) stanowi on (prawie) zawsze formułę niekompletną. Dopiero dobierając do określonego orzeczenia *B* jeden z możliwych podmiotów (np. *A* = „3”) otrzymujemy zamkniętą, możliwą do oceny pod kątem prawdziwości, semantyczną całość.

Mając na uwadze opisane prawidłowości syntaktyczne, mógł Frege sięgnąć do matematycznej teorii funkcji i za jej pomocą wyjaśnić formalnie, na czym polega znaczenie orzecznika (nazwy generalnej). W tym celu przeprowadził dwa zabiegi na raz: uogólniający i uszczegóławiający. Z jednej strony rozszerzył pojęcie argumentu funkcji (arytmetycznej), tak by stosowało się ono nie tylko do liczb, lecz do dowolnych obiektów mających nazwy w przyjętym języku. Z drugiej strony, spośród wszystkich możliwych funkcji wybrał *funkcje logiczne*¹⁵, czyli takie, których przeciwdziedzina jest zbiór binarny {Prawda, Fałsz}. Za ostatnią decyzją krył się

¹⁴ W przypadku pojęcia „pierwiastek kwadratowy czwórki” odpowiednia funkcja *p* mogłaby mieć postać $p(x) := (= (x \cdot x, 4))$. W jej definicji występują dwie inne funkcje: kwadratowa i tożsamościowa (przyjmująca wartość 1 tylko wtedy, gdy jej obydwa argumenty są tożsame).

¹⁵ Frege nie używał terminu „funkcja logiczna”. Operował dłuższym określeniem: „funkcja, której wartością dla każdego argumentu jest wartość logiczna” (Frege 1977, s. 28).

postulat ostrości pojęć, czyli stanowisko, zgodnie z którym wszelkie orzeczenia dopełnione podmiotami stają się bądź prawdziwe, bądź fałszywe¹⁶.

Metodologiczny zabieg utożsamienia pojęć z funkcjami logicznymi pozwolił Fregemu dostrzec jeszcze jedno ważne zagadnienie: wielość reprezentacji danego pojęcia. W przypadku funkcji liczbowych stanowi ona konsekwencję mnogości elementarnych operacji arytmetycznych, które mogą składać się na konkretną formułę funkcji. Na przykład wyrażenie $(2x)^2$ daje się zapisać jako $(x + x)^2$, $(x + x)(x + x)$ i na wiele innych, równoważnych sposobów. W przypadku pojęć wielość reprezentacji wynika z analogicznej mnogości elementarnych środków zapisu wiedzy, chociażby spójników logicznych. Z tego powodu pojęciem należy nazwać nie konkretną reprezentację danej funkcji logicznej, ale jej przebieg, czyli rozkład przyporządkowań wartości logicznych wielkościom z pewnej dziedziny. Postępując w ten sposób, pojęcia rozważa się abstrakcyjnie – abstrahując od konkretnych reprezentacji, zapisów itd. – a to sprawia, że można je badać za pomocą narzędzi matematyczno-logicznych.

Jeśli to, co powiedziano powyżej podsumujemy przy użyciu terminologii dostosowanej do potrzeb dalszych rozważań, otrzymamy taki oto obraz pojęć. Są one rozumiane ogólnie jako funkcje o określonej dziedzinie (np. dziedzinie obiektów fizycznych) i binarnym zbiorze wartości – co można zapisać symbolicznie jako $p: U \rightarrow \{0,1\}$. Zbiór $Z = \{x: p(x) = 1\}$ stanowi zakres pojęcia p , a konkretna formuła $p(x)$ to jego treść (przy czym formuł wyznaczających ten sam zakres może być wiele).

Podkreślmy na koniec, że dopiero na gruncie przygotowanym przez Fregego uwidacznia się różnica między pojęciami w sensie psychologicznym i pojęciami w sensie logicznym. Zgodnie z jego koncepcją o pojęciach w sensie logicznym wolno mówić dopiero wtedy, gdy są spełnione dwa warunki: **po pierwsze, abstrahuje się od konkretnych reprezentacji pojęć (np. wewnątrzmoźgowych)**¹⁷ **i przeżyć osób operujących pojęciami, a, po drugie, wynik procesu abstrakcji, czyli pewnego rodzaju funkcję, czyni się przedmiotem teorii formalnej (zmatematyzowanej)**. W teorii takiej należy określić przede wszystkim klasę

¹⁶ Przedstawiając zagadnienie niekompletności pojęć-funkcji w interpretacji psychologicznej (zob. podrozdział 1.4) – co może stanowić tylko pewną ilustrację czysto logicznego ujęcia Fregego – powiedzielibyśmy, że ludzie uaktywniają pojęcia (formy niekompletne) na dwa sposoby. Z jednej strony, „dopełniają je przedmiotami”, na przykład rozpoznają przedmioty (jako podpadające pod dane pojęcie), odróżniają przedmioty (należące do zakresów dwóch różnych pojęć) i określają własności przedmiotów (będących elementami treści odpowiednich pojęć). Z drugiej strony, wykorzystują pojęcia w postaci nieuzupełnionej, na przykład wtedy, gdy łączą dwa pojęcia w pojęcie złożone lub formułują ogólne prawa posługiwania się pojęciami.

¹⁷ Mówiąc o konkretnej reprezentacji pojęcia można mieć na myśli różne poziomy opisu tych reprezentacji. Na poziomie biologicznym czy fizjologicznym chodzi o pewne struktury neuronalne w mózgu konkretnej osoby – struktury odpowiedzialne za posługiwanie się pojęciem, w tym za rozpoznawanie jego desygnatów. Na poziomie psychologicznym chodzi o pewne schematyczne opisy zbiorów wyróżnionych obiektów, charakteryzujące zawartość pamięci podmiotu, a formułowane niekiedy przy użyciu języka informatyki, np. konkretnych struktur danych (takich jak drzewa decyzyjne czy kompleksy cech; por. rozdział 4).

funkcji odpowiadających pojęciom (a zatem wskazać ich dziedzinę, zbiór wartości i pewne ogólne własności, np. dyskretność lub ciągłość) oraz dopuszczalne działania na funkcjach wskazanej klasy. Po ich określeniu można przystąpić do formułowania i dowodzenia twierdzeń dotyczących wszelkich pojęć-funkcji, niezależnie od konkretnych postaci ich formuł. Sugestywnej analogii dostarcza tu rachunek różniczkowy: stanowi on ogólną teorię funkcji różniczkowalnych (a więc funkcji określonego typu), której twierdzenia dotyczą wszelkich takich funkcji, niezależnie od tego, jakimi wzorami się one wyrażają¹⁸.

2.2. Interpretacja operacjonistyczna koncepcji Fregego¹⁹

Wybrane przez Fregego matematyczne narzędzie opisu pojęć, a więc *funkcja*, ma z filozoficznego punktu widzenia tę zaletę, iż nie zobowiązuje ontologicznie. Posługujący się nim teoretycy nie muszą zakładać istnienia obiektów odpowiadających funkcjom. Kwestię tę obrazuje dobrze analiza zdań typu „*Y* jest funkcją *X*-a”, np. „Droga jest funkcją czasu”. Otóż w zdaniach takich termin „funkcja” pełni rolę skrótowo-zastępczą; sprawia, że znaczą one tyle, co „*Y* pozostaje w zależności od *X*-a” lub dokładniej „Gdy zmienia się *X*, zmienia się *Y*” (np. „Wraz z upływem czasu, droga (przebyta przez jakieś ciało) zmienia się tak a tak”).

Podążając konsekwentnie za myślą Fregego, musimy przyjąć, iż termin „pojęcie” wykazuje istotne podobieństwo semantyczne do terminu „funkcja” (za pomocą którego wszak go się objaśnia). Z tego względu musi pełnić rolę skrótowo-zastępczą: ani nie oznacza niczego, ani nie odnosi w żaden sposób do przeżyć osoby dysponującej pojęciem. Sprawę tę wyjaśnią następujące, przykładowe tłumaczenia zdań typu „*X* ma pojęcie *M*-a” – tłumaczenia, spośród których wybierzemy to, które naszym zdaniem najlepiej oddaje myśl Fregego. Oto one:

- a) *X* rozumie nazwę oznaczającą *M*-y, np. nazwę „*M*”,
- b) *X* dostrzega podobieństwo między *M*-ami,
- c) *X* klasyfikuje obiekty z pewnej dziedziny jako *M*-y i nie *M*-y,
- d) W pamięci *X*-a są reprezentowane *M*-y (odróżnione od nie *M*-ów),

¹⁸ Dopowiedzmy tu, że Frege, jakkolwiek zarysował program abstrakcyjnego ujmowania pojęć jako funkcji, w swoich badaniach skupiał się raczej na formalnej teorii zdań, w których pojęcia występują. Teoria taka miała na celu wypracowanie metod określania prawdziwości zdań złożonych na podstawie prawdziwości zdań prostych. W tym sensie Fregego należy uznać raczej za prekursora rachunku zdań (odpowiadających sądom) niż nazw (odpowiadających pojęciom).

¹⁹ Chodzi o interpretację autorską, która naszym zdaniem jest zgodna z abstrakcyjnym rysem koncepcji Fregego, ale też pozwala zrozumieć lepiej, na czym polega specyfika współczesnego informatycznego podejścia do pojęć, w ramach którego – podobnie jak u Fregego – pojęcia utożsamia się z pewnymi funkcjami.

- e) *W mózgu X-a istnieje struktura neuronalna odpowiedzialna za rozpoznawanie M-ów.*

Zaproponowane przekłady wymagają krótkiego komentarza. W zdaniach a), b) i c) termin „pojęcie” ma charakter skrótowo-zastępczy; nie oznacza żadnego obiektu fizycznego, lecz informuje o tym, że *X* coś robi (rozumie, dostrzega podobieństwo lub klasyfikuje) w odpowiedzi na takie a nie inne okoliczności, np. w obecności pewnego obiektu i/lub pewnej nazwy. W odróżnieniu od przekładu c), w którym mówi się o czynności dostępnej obserwatorom zewnętrznym (klasyfikowaniu), w tłumaczeniach a) i b) wychodzi na jaw założony mentalizm, to znaczy fakt, że o posiadaniu przez kogoś pojęcia przesądza tak naprawdę jego subiektywne przeżycia (rozumienie i dostrzeganie podobieństwa). W zdaniach d) i e) termin „pojęcie” ma charakter dosłowny, a nie skrótowo-zastępczy; oznacza bowiem pewien obiekt: fizyczny (struktura neuronalna) lub teoretyczny (reprezentacja poznawcza).

Naszym zdaniem to tłumaczenie c) – nie odwołujące się ani do mentalnej, ani do fizycznej strony pojęć – najlepiej oddaje zapoczątkowaną przez Fregego tendencję do opisu pojęć za pomocą funkcji. Przyjmując zasadność takiego sposobu opisu, zdanie c) można przekształcić dalej tak, by termin „pojęcie” ustąpił miejsca terminowi „funkcja”. Otrzymujemy wówczas taki oto przekład zdania „*X* ma pojęcie *M*-a”: „Decyzje klasyfikacyjne *X*-a są funkcją klasyfikowanych obiektów” lub – zważywszy na zastępczy charakter wyrazu funkcja – jeszcze krócej „Gdy zmieniają się obiekty, zmieniają się decyzje”.

Zabieg, którego dokonaliśmy – będący niewątpliwie pewną interpretacją – pozwala nam uznać podejście Fregego za czysto *operacyjne*. **Pojęcia-funkcje służą tu do czysto abstrakcyjnego opisu możliwych zachowań (decyzji) kategori-zacyjnych pewnych systemów** (zarówno naturalnych, np. ludzi i zwierząt, jak i sztucznych, np. systemów sztucznej inteligencji)²⁰. Z tego powodu pojęcia w rozumieniu Fregego, nazywane wcześniej funkcjami logicznymi, wolno opatrzyć etykietą funkcji decyzyjnych.

Definiowanie pojęć jako funkcji decyzyjnych przyjęło się w informatyce, gdzie abstrakcyjny, matematyczny opis pojęć-funkcji jest wykorzystywany np. do analizy problemów rozpoznawania obiektów i generowania złożonych klasyfikacji obiektów (Cichosz 2000; Flasiński 2011). W informatyce, a zwłaszcza w badaniach nad sztuczną inteligencją, istotne są jednak nie tylko czysto matematyczne rozważania o typach i własnościach odpowiednich funkcji, ale też ich

²⁰ Za operacyjną interpretacją koncepcji Fregego przemawia również fakt, że na zbiór wartości funkcji odpowiadających pojęciom wybrał zbiór $\{0,1\}$, a nie rodzinę wszystkich podzbiorów uniwersum, 2^U (pojęcia to funkcje $f: U \rightarrow \{0,1\}$, a nie $f: U \rightarrow 2^U$). Naszym zdaniem, wybór taki uwypukla myśl, iż wykorzystywanie pojęć polega na podejmowaniu decyzji (np. percepcyjnych), a nie na przywoływaniu całego zakresu pojęcia. Mówiąc swobodnie: „mieć pojęcie”, to tyle co „umieć je wykorzystać w różnych sytuacjach”, a nie – „umieć podać jego zakres”.

szczególnego rodzaju interpretacje (będące informatycznymi odpowiednikami wzorów funkcji). Reprezentacje te dobiera się odpowiednio do typu systemu i zadań, które system realizuje. Przykładowo, w systemach opartych na sztucznych sieciach neuronowych, które służą do rozpoznawania obiektów, stosuje się reprezentacje koneksyjne (są to układy wag między sztucznymi neuronami sieci) o strukturze zgodnej z topologią sieci. Kwestie te omówimy szerzej w rozdziale 4.

Podejście informatyczne jest zatem pewnym uszczegółowieniem koncepcji Fregego, w ramach którego można badać w sposób formalny określone typy reprezentacji. Jest to ujęcie intensjonalne (a nie czysto ekstensjonalne), ponieważ bierze się w nim pod uwagę postać funkcji odpowiadających pojęciom, a także zależną od tej postaci sprawność systemu (np. dla pewnych reprezentacji mamy do dyspozycji szybsze algorytmy, dla innych wolniejsze). Mimo wszystko ostateczne kryterium przydatności danej reprezentacji, to poprawność podejmowanych przez system decyzji klasyfikacyjnych, a więc kryterium operacyjne²¹.

2.3. Możliwe rozszerzenia koncepcji Fregego

Centralnym elementem koncepcji Fregego jest postulat ostrości, zgodnie z którym pojęcia w sensie logicznym (znaczenia nazw generalnych) powinny być opisywane za pomocą funkcji przyporządkowujących obiektom z pewnego uniwersum U ostre i zero-jedynkowe decyzje kategoryzacyjne, czyli funkcji typu $U \rightarrow \{0,1\}$. Postulat ten **można rozszerzać na różne sposoby**, zachowując przy tym podejście abstrakcyjno-formalne, abstrahując od dokładnych kształtów formuł pojęć-funkcji oraz od indywidualnych charakterystyk (np. przeżyć) „posiadaczy pojęć”. Rozszerzenia te będą omawiane w kolejnym rozdziale książki przy okazji analizy związków między teoriami pojęć i teoriami zbiorów. W tym podrozdziale poczynimy jednak pewne wstępne obserwacje.

Przede wszystkim binarny zbiór wartości pojęcia-funkcji można zastąpić zbiorem obszerniejszym, choć ograniczonym z dołu i z góry liczbami 0 i 1. W ten sposób zostaje odrzucony fregowski postulat ostrości pojęć. Pojęcia przyjmują postać funkcji $U \rightarrow D$, gdzie D oznacza dowolny podzbiór przedziału $[0,1]$ zawierający 0 i 1 (jako ograniczenia dolne i górne)²². Przyjmując interpretację operacyjną pojęcia, powiemy, że decyzję o podpadaniu obiektów uniwersum pod pojęcie

²¹ Chociaż z punktu widzenia problemu rozważanego w ostatnim rozdziale książki, problemu czarnej skrzynki, istotne jest także kryterium nieoperacyjne, związane z przejrzystością danej metody reprezentacji dla człowieka (gdzie przejrzystość rozumiemy jako zdolność do generowania przez system satysfakcjonujących człowieka wyjaśnień podejmowanych decyzji).

²² Jako przykładowe warianty zbioru $D \neq \{0,1\}$ można wymienić: a) D skończony (np. $\{0, 1/2, 1\}$), b) D nieskończony i niegęsty (np. zawierający liczby postaci $1/n$), c) D nieskończony i gęsty (zawierający wszystkie liczby rzeczywiste z przedziału $[0,1]$).

wykraczają poza jednoznaczne odpowiedzi TAK lub NIE, bo z każdą decyzją jest związany pewien stopień pewności. Podstawą rachunku pojęć o zakresach nieostrych staje się wówczas pewna logika wielowartościowa i odpowiadający jej wariant *teorii zbiorów rozmytych*²³ (Zadeh 1975). Kwestię tę omówimy szczegółowo w podrozdziale 3.2.

Analizując głębiej zagadnienie pochodzenia nieostrości pojęć, można postawić pytanie o możliwość określenia pojęcia-funkcji p na podstawie pojęć ostrych $p_1, \dots, p_n$, składających się na wiedzę W o obiektach uniwersum (wiedzę pewną). Pytanie to dotyczy precyzji opisu p na podstawie W . Może się zdarzyć, że zakres pojęcia p daje się określić jedynie w przybliżeniu, za pomocą tzw. przybliżenia górnego i dolnego, a tym samym na gruncie wiedzy W jest niepewny. To zaś sprawia, że zakres odpowiedniego pojęcia, o ile chcielibyśmy go zdefiniować w oparciu o wiedzę W , jest nieostry. Podstawę dla takich rozważań stanowi *teoria zbiorów przybliżonych* (Pawlak 1991), której związkom z teorią pojęć przyjrzymy się w podrozdziale 3.3.

Kolejna możliwość rozwinięcia myśli Fregego polega na bliższym wniknięciu w postać funkcji przyporządkowujących obiektom decyzje klasyfikacyjne (ostre lub nieostre). Chodzi tu o spostrzeżenie ogólne, tj. nadal abstrahujące od kształtu formuły funkcji, że wiele (być może większość) pojęć definiuje się za pośrednictwem wyróżnionych cech obiektów. Z tego względu funkcję $p: U \rightarrow D$ należy traktować jako superpozycję dwóch innych funkcji, c i d . Pierwsza z nich przekształca obiekty na ich opisy w kategoriach cech ($c: U \rightarrow C_1 \times C_2 \times \dots \times C_n$, gdzie C_i jest zbiorem wartości i -tej cechy), a druga przyporządkowuje kombinacjom wartości cech odpowiednie decyzje kategoryzacyjne ($d: C_1 \times C_2 \times \dots \times C_n \rightarrow D$). Ujęcie takie przyjęło się, na przykład, w informatyce (zob. (Cichosz 2000)) i będziemy do niego nawiązywać w rozdziale 4.

Zagadnienie, o którym wspomniemy na koniec, dotyczy dziedziny funkcji odpowiadających pojęciom. Otóż Frege stawia je tak, jak gdyby przyjmował istnienie *wyidealizowanego języka*, w którym każdemu pojęciu odpowiada dokładnie jedna nazwa (poszczególne nazwy są reprezentowane przez poszczególne obiekty uniwersum U). W konsekwencji różne *wypowiedzenia* danej nazwy – różne co do kształtu czy brzmienia, ale odpowiadające temu samemu pojęciu – traktuje jako nierozróżnialne. Tymczasem rzeczywiste decyzje co do przynależności obiektu do zakresu pojęcia są podejmowane zawsze w „obecności” konkretnego wypowiedzenia danej nazwy, np. w pewnym kontekście sytuacyjnym, który wpływa istotnie na sposób rozumienia wypowiedzi. Można zastanawiać się, czy w abstrakcyjnej teorii pojęć nie należy tego faktu uwzględnić. Po jego uwzględnieniu pojęcie byłoby funkcją $W_N \times U \rightarrow \{0,1\}$, gdzie W_N stanowi uniwersum

²³ Pisząc o wariacie teorii zbiorów rozmytych, mamy na myśli pewien zestaw definicji podstawowych operacji teoriomnościowych (jak suma czy przecięcie zbiorów), który definiuje określony wariant (Klir, Yuan 1995).

wypowiedzeń nazw. Temat ten nie będzie rozwijany w dalszej części książki, ale warto zauważyć, że w realnych sytuacjach komunikacyjnych, to właśnie kontekst wypowiedzi sprawia, że pojęcia mają charakter nieostry. Przykładowo, termin „wysoki” będzie rozumiany inaczej w rozmowie koszykarzy, a inaczej w rozmowie dżokejów. Próbując wymodelować znaczenie tego terminu z uwzględnieniem różnych możliwych kontekstów, możemy potraktować go jako termin nieostry i sięgnąć do teorii zbiorów rozmytych (zob. podrozdział 3.2)²⁴.

2.4. Podsumowanie

1. Ze względu na fakt, że pojęcia określają przyporządkowania obiektom z pewnego uniwersum ostrych lub nieostrych decyzji o przynależności obiektów do zakresów nazw, najlepszym matematycznym narzędziem opisu pojęć są **funkcje** (w tym: binarne funkcje decyzyjne).

2. Zgodnie z tradycją zapoczątkowaną przez G. Fregego o pojęciach w sensie logicznym wolno mówić, gdy są spełnione dwa warunki: a) **abstrahuje** się od wewnątrzrozumysłowych reprezentacji pojęć i przeżyć osób operujących pojęciami, b) wynik procesu abstrakcji, czyli funkcję, czyni się przedmiotem **teorii formalnej** (zmatematyzowanej). Warunki te spełnia oryginalna koncepcja Fregego, w której pojęcia rozumie się jako binarne funkcje decyzyjne.

3. Zgodnie z **operacjonistyczną** interpretacją ujęcia fregowskiego i ujęć postfregowskich pojęcia-funkcje (mówiąc dokładniej: funkcje decyzyjne) służą do czyisto **abstrakcyjnego** opisu możliwych decyzji kategoryzacyjnych ludzi (a szerzej: systemów klasyfikujących, również sztucznych).

4. Oryginalną koncepcję Fregego **rozszerza się** współcześnie w trzech kierunkach: a) rezygnuje się z postulatu **ostrości** pojęć, dopuszczając rozmyte zakresy pojęć; b) proponuje się różnego rodzaju formalne **reprezentacje** pojęć, odwołujące się do cech desygnatów; c) wprowadza się dodatkowe miary podobieństwa między pojęciami należącymi do pewnej wspólnej przestrzeni.

²⁴ Jeszcze innym rozwinięciem koncepcji Fregego jest teoria przestrzeni pojęciowych (*conceptual spaces*) Petera Gardenforsa, w ramach której, po pierwsze, wnika się w kształty formuł definiujących pojęcia-funkcje na podstawie wyróżnionych uprzednio cech (jest to zatem ujęcie intensjonalne), po drugie zaś, określa się pewną metrykę (na wzór analizy matematycznej) pozwalającą wyznaczać „odległości między pojęciami”, czyli stopnie ich podobieństwa. Zob. (Gärdenfors 2004).

Rozdział 3

Teorie pojęć a teorie zbiorów

Zgodnie z tym, co sugeruje powyższy tytuł, we współczesnej nauce (zwłaszcza w obrębie podstaw matematyki) mamy do czynienia z różnymi teoriami zbiorów. I nie chodzi tu tylko o to, że w ramach pewnej koncepcji, na przykład w klasycznej teorii mnogości, przyjmuje się różne terminy podstawowe i/lub różne warianty aksjomatyzacji. Czasami rozbieżności sięgają głębiej, a dotyczą bądź różnego rozumienia terminu *zbiór*, bądź różnych założeń wstępnych, co do przynależności elementów do zbioru.

W dalszej części rozdziału zagadnienie powyższe rozważymy na przykładzie dwóch szczególnie wyrazistych przeciwstawień. Z jednej strony, pozostając przy tradycyjnym czyli dystrybutywnym sposobie rozumienia terminu „zbiór”, postawimy naprzeciw siebie zbiory *ostre*, a więc grupy elementów wyróżnionych z ogółu z całą pewnością (np. ze względu na jednoznaczne kryteria), i zbiory *rozmyte*, a więc grupy elementów wyróżnionych z ogółu mniej lub bardziej stanowczo²⁵. Z drugiej strony, wnikając głębiej w znaczenie terminu „zbiór”, rozróżnimy zbiory *kolektywne*, czyli całości złożone z części, i zbiory *dystrybutywne*, czyli grupy odrębnych elementów, wyróżnianych ze względu na pewne ich własności. Obydwie dystynkcje omówimy pod kątem formalnego opisu pojęć (mając oczywiście na uwadze, że ich zakresami są zawsze pewne zbiory). Wcześniej jednak przyjrzymy się bliżej temu, jak opis pojęć w ramach takiej czy innej teorii zbiorów wiąże się z właściwą pojęciom, zwłaszcza przy podejściu psychologicznym, funkcją reprezentowania.

3.1. Zakresy i formuły pojęć-funkcji

W poprzednim rozdziale wielokrotnie podkreślaliśmy zasadniczą różnicę między *logicznym* i *psychologicznym* sposobem opisu pojęć. Sposób pierwszy polega na

²⁵ Jak na razie teorię zbiorów rozmytych sformułowano dla zbiorów rozumianych dystrybutywnie. Nic jednak nie stoi na przeszkodzie, aby opracować jej odpowiednik dla zbiorów kolektywnych.

włączeniu pojęć w pewną abstrakcyjną teorię funkcji decyzyjnych; abstrakcyjną w tym sensie, iż nie uwzględnia się w niej ani formuł owych funkcji, ani przeżyć osób posługujących się nimi. W ujęciu drugim czyni się przeciwnie: zakłada się przynależność pojęć do umysłu podmiotu i bada się co najmniej właściwe im funkcje reprezentowania (coś reprezentują) i pośredniczenia (pośredniczą między postrzeganymi obiektami i decyzjami podmiotu). Zważywszy na powyższe rozróżnienia, możemy stwierdzić, co następuje.

Po pierwsze, **abstrakcyjne teorie pojęć-funkcji są tożsame z teoriami zbiorów w sensie dystrybutywnym**²⁶. Teza ta wynika wprost z faktu, iż jedna z metod określania zbiorów – w przypadku zbiorów nieskończonych jest to metoda jedyna – polega na definiowaniu ich *funkcji charakterystycznych*. Ujmując rzecz ściśle: każdemu zbiorowi ostremu A odpowiada funkcja charakterystyczna $f_A: X \rightarrow \{0,1\}$, taka że $f_A(x) = 1$ wtedy i tylko wtedy, gdy $x \in A$, oraz $f_A(x) = 0$ w przeciwnym wypadku; natomiast każdemu zbiorowi rozmytemu Z odpowiada funkcja charakterystyczna $f_Z: X \rightarrow [0,1]$, taka że $f_Z(x) = 1$, gdy $x \in Z$ z całą pewnością, $f_Z(x) \in (0,1)$, gdy $x \in Z$ w pewnym tylko stopniu oraz $f_Z(x) = 0$, gdy x z całą pewnością nie należy do Z . Jak widać więc, każdy zbiór (ostry lub rozmyty) można utożsamić z jego funkcją charakterystyczną.

Jeśli przyjmiemy zatem, iż teoria pojęć w sensie logicznym jest teorią wszelkich funkcji decyzyjnych określonego rodzaju (teorią, w której nie uwzględnia się postaci formuł funkcyjnych), to musimy zgodzić się, iż pojęcia – będące funkcjami charakterystycznymi zbiorów – są tożsame ze swoimi zakresami. Innymi słowy, musimy przyjąć, że wszystko to, co daje się ustalić odnośnie pojęć-funkcji, tak naprawdę dotyczy ich zakresów (niekoniecznie ostrych).

Po drugie jednak, gdy pojęcia są rozpatrywane na niższym poziomie abstrakcji, rzecz wygląda zgoła inaczej. Wtedy bowiem liczą się nie tylko dziedziny i przeciwdziedziny pojęć-funkcji, ale również kształt ich formuł. Z tego względu niektóre przynajmniej pojęcia równozakresowe uznaje się za różne, a w konsekwencji teorie pojęć stają się czymś więcej niż teoriami zakresów pojęć (czyli zbiorów). Zagadnienie to rysuje się szczególnie wyraziście na gruncie psychologii, gdzie **formułom pojęć-funkcji nadaje się znaczenie reprezentujące**. Uznaje się zatem, że formuły te pośredniczą między umysłem i światem, dokładniej zaś reprezentują w umyśle, w określony sposób, zbiory obiektów wyróżnionych w świecie²⁷.

Teorie zbiorów stanowią wówczas formalne narzędzie obiektywnej (logicznej) analizy świata reprezentowanego. Wybór teorii danego typu jest równoznaczny z przyjęciem pewnego wzorca tej analizy, określeniem tego, jak w idealnym

²⁶ Przy przyjętym w niniejszej pracy sposobie rozumienia pojęć w sensie logicznym.

²⁷ Na polu psychologii przyjmuje się dodatkowo, że istnieją pewne realne odpowiedniki formuł pojęć-funkcji, czyli formuł $p(x)$ – na przykład, reprezentacje poznawcze czy wręcz pewne struktury neuronalne.

wypadku powinna przebiegać kategoryzacja świata przez umysł. Na przykład, wybierając teorię zbiorów ostrych, zakłada się, że kategorie obiektów wyodrębnialnych w świecie mają ostre granice, a tym samym przyjmuje się, że w idealnej sytuacji percepcyjnej możliwe jest jednoznaczne przyporządkowanie obiektów do określonych kategorii. Decydując się na teorię zbiorów rozmytych, odrzuca się powyższe założenie. Reasumując, można powiedzieć, że psychologiczne teorie pojęć nie są tożsame z formalnymi teoriami zbiorów, lecz są z nimi normatywnie skorelowane.

W jednym z tekstów *stricte* psychologicznych znajdujemy takie oto wyjaśnienie tej kwestii: „*Podstawowym założeniem w psychologicznych koncepcjach pojęć jest rodzaj formalnej teorii zbiorów przyjmowany za punkt wyjścia. Typowym sposobem realizacji tego założenia w psychologicznej teorii pojęć jest: (1) przyjęcie, że na poziomie świata reprezentowanego dany typ teorii zbiorów jest właściwym narzędziem „obiektywnej” analizy struktury kategorialnej tego świata, (2) przyjęcie, że formalna struktura pojęć reprezentujących owe kategorie jest w zasadzie zgodna ze strukturą tej właśnie „obiektywnej” analizy*” (Najder 1993, s.153).

W powyższych uwagach wstępnych usiłowaliśmy wyjaśnić, iż zależnie od stopnia ogólności dyskusji, teorie pojęć są bądź tożsame, bądź częściowo tożsame, bądź normatywnie skorelowane z formalnymi teoriami zbiorów. Jednocześnie informowaliśmy o wielości tych ostatnich. I tutaj właśnie rodzi się pytanie: czy różne formalne teorie zbiorów stanowią konkurujące ze sobą narzędzia opisu wszelkich pojęć, czy też jest odwrotnie – są one użyteczne przy analizie pojęć różnych rodzajów. W kolejnych podrozdziałach postaramy się pokazać, że ma miejsce raczej to drugie.

3.2. Pojęcia ostre i rozmyte

Pojęciom ostrym, czyli klasycznym, poświęciliśmy większą część rozdziału 2, w którym przedstawiliśmy zinterpretowaną współcześnie koncepcję Fregego. Z metodologicznego punktu widzenia pojęcia takie są wprowadzane do języka za pomocą precyzyjnych definicji, które określają jednoznacznie zbiory obiektów należące do zakresu danego pojęcia. Metodologicznym wzorcem są tutaj definicje formułowane w ramach nauk ścisłych. Formalnie rzecz biorąc, **pojęcia ostre są binarnymi funkcjami decyzyjnymi**, czyli takimi funkcjami, które przyjmują tylko dwie wartości: 0 i 1. Wartość 1 informuje, że dany obiekt należy do zakresu pojęcia, wartość 0 – że nie należy. Owe binarne funkcje decyzyjne są jednocześnie funkcjami charakterystycznymi zbiorów, zbiorów ostrych, które stanowią zakresy pojęć.

Podstawę ścisłego opisu pojęć nieostrych, czyli nieklasycznych, stanowi teoria zbiorów rozmytych (TZR). Teoria ta wywodzi się bezpośrednio z logik

wielowartościowych, a więc takich logik, w których dopuszcza się ocenę zdań za pomocą innych wartości logicznych niż „prawda” i „fałsz” (Bolc 1992). Jedną z przyczyn ich opracowywania jest **występowanie w języku naturalnym słów i zwrotów, które nie wyznaczają swojego zakresu w sposób jednoznaczny**. W najprostszym wypadku odpowiadają one własnościom, które przysługują desygnatom z różnymi stopniami nasilenia lub mogą być przypisywane im z różnymi stopniami pewności. Do typowych przykładów wyrażeń nieostrych zaliczają się: *młody*, *wysoki*, *dużo większy niż* (np. cecha *młody* w innym stopniu przysługuje piętnastolatkowi, a w innym trzydziestolatkowi). Ich znaczenia modeluje się właśnie za pomocą zbiorów rozmytych²⁸.

Formalnie rzecz biorąc, zbiorem rozmytym A nad pewnym uniwersum X nazywa się zbiór par $(x, m_A(x))$, gdzie x jest elementem uniwersum X , a $m_A: X \rightarrow [0,1]$ funkcją charakterystyczną zbioru A , nazywaną też funkcją przynależności elementów X do zbioru A . Klasę wszystkich zbiorów rozmytych nad X oznacza się jako $F(X)$. W przypadku uniwersów o skończonej liczbie elementów ($X = \{x_1, \dots, x_n\}$) zbiór rozmyty zapisuje się jako zbiór par uporządkowanych, $\{(x_1, m_A(x_1)), \dots, (x_n, m_A(x_n))\}$, zaś w przypadku uniwersów ciągłych podaje się odpowiednią formułę funkcji przynależności, np. $m_A(x) = 1 - x^2$ dla $x \in [-1,1]$.

Definicja zbioru rozmytego jako zbioru par stanowi uogólnienie klasycznej koncepcji zbioru. Funkcje przynależności zbiorów klasycznych przyjmują bowiem tylko wartości 0 (element nie należy do zbioru) i 1 (element należy do zbioru)²⁹; tutaj zaś dopuszcza się dowolne wartości z przedziału $[0,1]$. Przykładowo: dla zbioru A odpowiadającego nazwie „młody” i określonego na uniwersum liczb naturalnych (lat życia) od 10 do 50, $m_A(20)$ mogłoby wynosić 1, zaś $m_A(40)$ mogłoby być równe 0,7.

Podstawowe działania na zbiorach rozmytych, takie jak suma, przecięcie i dopełnienie, stanowią uogólnienia odpowiednich operacji na zbiorach ostrych; ich definicje są dobrane tak, by w przypadku zbiorów o zero-jedynkowych funkcjach charakterystycznych uzyskać zgodność z definicjami dla zbiorów ostrych. W efekcie dokonania na zbiorze lub zbiorach pewnej operacji powstaje wynikowy zbiór W o pewnej funkcji przynależności m_W . Istota operacji kryje się w metodzie wyznaczania m_W na podstawie danych funkcji m_A i m_B . Mówiąc intuicyjnie, metoda ta musi dostarczyć informacji o tym, „na ile” poszczególne elementy uniwersum przynależą do zbioru wynikowego, jeśli wiadomo „na ile” przynależą one do zbiorów A i B .

²⁸ Podstawowe idee teorii zbiorów rozmytych można znaleźć w pracy twórcy tej teorii (Zadeh 1975), zaś ich poszerzone opracowanie zawarte są w pracy (Folger, Klir 1988).

²⁹ Z tego powodu funkcję przynależności się pomija, a zbiór określa się, wypisując po prostu te elementy, dla których wartością funkcji przynależności jest 1 (ewentualnie: podając własność, dla której można jednoznacznie wskazać mające ją elementy).

Tradycyjne definicje *sumy*, *przecięcia* i *dopełnienia* pochodzą od twórcy teorii zbiorów rozmytych, Lofti A. Zadeha. Wypiszmy kolejno: $m_{A \cup B}(x) = \min[m_A(x), m_B(x)]$, $m_{A \cap B}(x) = \max[m_A(x), m_B(x)]$, $m_{A'}(x) = 1 - m_A(x)$, gdzie $A, B, A' \in F(X)$, $x \in X$ ³⁰.

Kolejna, ważna z punktu widzenia semiotyki, klasa działań na zbiorach rozmytych to ich przekształcenia przy użyciu *modyfikatorów lingwistycznych*. Za ich pomocą modeluje się znaczenia funktorów nazwotwórczych od argumentu nazwowego, w szczególności takich jak *bardzo* i *raczej*. Definicja modyfikatora związanego z funktorem f informuje o tym, jaki zbiór rozmyty ma odpowiadać wyrażeniu $f(w)$, jeśli wiadomo, jaki zbiór skojarzono z wyrażeniem w . Na przykład, jeśli wyrażeniu *wysoki* odpowiada pewien zbiór $A \in F([0, 250])$, to znaczenie zwrotu *bardzo wysoki* oddaje zbiór $\text{bardzo}(A) \in F([0, 250])$. Typowe definicje modyfikatorów *bardzo* i *raczej* wyglądają następująco: (1) jeśli $C = \text{bardzo}(A)$, to $m_C(x) = (m_A(x))^2$, (2) jeśli $C = \text{raczej}(A)$, to $m_C(x) = (m_A(x))^{0.5}$.

Rozwinięciem koncepcji modyfikatorów rozmytych jest pojęcie *zmiennej lingwistycznej*. Zmienna taka przebiega zbiór wyrażen budowanych z pewnych terminów pierwotnych przy użyciu ściśle określonych modyfikatorów rozmytych. Mówiąc dokładniej, zmienną definiują: (1) uniwersum X , (2) zestaw terminów pierwotnych (np. *niski*, *wysoki*, *średni*), (3) zestaw zbiorów rozmytych nad uniwersum X odpowiadających poszczególnym terminom pierwotnym, (4) zestaw modyfikatorów rozmytych odpowiadających za przekształcanie zbiorów z $F(X)$, (5) zestaw reguł budowania terminów złożonych z terminów pierwotnych (np. *bardzo niski* czy *raczej niewysoki*). Jak widać więc, wartościami zmiennej lingwistycznej są pewne etykiety słowne, rozumiane nieostro, dotyczące pewnej wspólnej wielkości (np. wzrostu) i budowane z terminów pierwotnych.

W zastosowaniach teorii zbiorów rozmytych podstawową rolę odgrywa *operator przypisania*, za pomocą którego tworzy się *dane rozmyte*. Dane takie należy traktować jako wiązki czy paczki danych ostrych z dołączonymi doń stopniami pewności. Formalnie fakt utworzenia danej rozmytej zapisuje się jako „ x jest A ” (ang. „ x is A ”), gdzie zmienna x przybiera wartości z pewnego uniwersum X , a zbiór $A \in F(X)$. Wyjaśnienie to sugeruje, iż operator przypisania spełnia dwie funkcje: z jednej strony wskazuje uniwersum wartości pewnej zmiennej, a z drugiej – wiąże z każdą jej potencjalną wartością liczbę z przedziału $[0, 1]$. Przykładowe przypisania dotyczą werbalnych, nieostrych charakterystyk pewnych

³⁰ W rozszerzeniu oryginalnej koncepcji Zadeha zamiast tradycyjnych operatorów $\min$ i $\max$ stosuje się ich uogólnienia: specjalne funkcje $f: [0, 1] \times [0, 1] \rightarrow [0, 1]$ nazywane normami trójkątnymi. Definiuje się je aksjomatycznie. Proponuje się też metodę ich generowania z funkcji jednej zmiennej. Zdefiniowane za ich pomocą działania na zbiorach rozmytych mają postać: $m_{A \cap B}(x) = T[m_A(x), m_B(x)]$, $m_{A \cup B}(x) = S[m_A(x), m_B(x)]$, gdzie T i S są sprzężonymi ze sobą normami minimum i maksimum. Przykładami sprzężonych ze sobą par T i S są: (1) $T(a, b) = \max(0, a + b - 1)$, $S(a, b) = \min(1, a + b)$, oraz (2) $T(a, b) = ab$, $S(a, b) = a + b - ab$. Por. (Czogała, Pedrycz 1985).

Można powiedzieć, że wybór pary sprzężonych operatorów T i S , odpowiedzialnych za definicje sumy i przecięcia, wyznacza pewien wariant teorii zbiorów rozmytych.

wielkości liczbowych, np. „ x jest *Małe*”, „ x jest *Średnie*”, „ x jest *Duże*”, gdzie x jest liczbą rzeczywistą z przedziału $X = [0, 20]$, a *Małe*, *Średnie* i *Duże* są zbiorami rozmytymi nad uniwersum X .

Dane rozmyte najczęściej tworzy się po to, by określić numerycznie pewien nieostry, ale zrozumiały na gruncie języka naturalnego punkt wyjścia wnioskowania. Reprezentacja numeryczna sprawia, że odpowiednie dane można zapisywać w pamięci komputerów, a odpowiednie wnioskowania przeprowadzać algorytmicznie. W efekcie wnioskowania uzyskuje się inne dane rozmyte, które bądź pozostawia się w postaci nieostrej (dopasowując do nich pewną etykietę słowną), bądź się „wyostrza” (to znaczy wybiera się jedną, najbardziej reprezentatywną wartość zmiennej).

Teoria zbiorów rozmytych stanowi matematyczne narzędzie formalne, które daje się interpretować w różnych dziedzinach problemowych. Przykłady przytaczane w poprzednim ustępie sugerują interpretację polegającą na modelowaniu znaczeń terminów nieostrych, czyli modelowaniu pojęć nieostrych³¹. Wyrażając tę sugestię abstrakcyjnie, powiedzielibyśmy, że **dzięki teorii zbiorów rozmytych definicję pojęcia jako funkcji typu $f: U \rightarrow \{0,1\}$ daje się uogólnić do postaci $f: U \rightarrow [0,1]$** . W ten sposób zakres pojęcia staje się zbiorem rozmytym³².

Tutaj jednak pojawiają się dwa ważne pytania. Po pierwsze: „Czy próby modelowania pojęć nieostrych nie powinny ustąpić miejsca próbom ich uściślenia?”. I po drugie: „Nawet jeśli odrzucimy konieczność uściślenia, to jak należy rozumieć wartości funkcji przynależności do zbioru będącego zakresem pojęcia nieostrego?”.

Na negatywną odpowiedź na pytanie pierwsze musimy się zgodzić wówczas, gdy uwypuklamy psychologiczny aspekt pojęć, związany z tym, jak faktycznie ludzie formułują myśli i posługują się językiem. W języku potocznym zaś, a także przy próbach wstępnego opisu zagadnień naukowych, występują nader często terminy nieostre. Co więcej są one dla ludzi bardziej naturalne i komunikacyjnie

³¹ Wśród wielu interpretacji/zastosowań teorii zbiorów rozmytych znajdują się i takie, które z modelowaniem znaczenia słów nie mają wiele wspólnego. Jedno z popularnych zastosowań polega na rozwiązywaniu problemów diagnostycznych. Oto jest dana relacja R wiążąca ze sobą możliwe symptomy z potencjalnymi diagnozami, przy czym do każdego elementu tej relacji jest dołączona siła (w skali od 0 do 1) połączenia między danym symptomem a konkretną diagnozą (np. [(katar, angina); 0,8]). Relacja taka jest traktowana jako rozmyta. Symptomy i diagnozy są reprezentowane zbiorczo jako zbiory rozmyte – ich funkcje przynależności określają stopnie nasilenia poszczególnych objawów/diagnoz (np. {(katar; 0,3), (kaszel; 1), (wymioty; 0,7)}). Problem polega na takim wykorzystaniu relacji rozmytej R , aby dla różnych zestawów symptomów otrzymywać właściwe diagnozy. Próbuje się go rozwiązać za pomocą operacji złożenia relacji z rozmytym zbiorem symptomów, a następnie „wyostrzenia” uzyskanego wyniku, tj. wskazania jednego, najbardziej prawdopodobnego elementu z uniwersum diagnoz. Por. (Stacewicz 1994).

³² Taka definicja uogólniona stanowi podstawę psychologicznych, prototypowych koncepcji reprezentacji pojęciowych. W różnych koncepcjach tego typu proponuje się różne interpretacje funkcji przynależności (różne metody ich określania). Por. np. (Smith, Medin 1981).

skuteczne niż precyzyjne definicje czy suche dane liczbowe. Fakty te są szczególnie istotne z perspektywy badań nad sztuczną inteligencją, których jednym z celów jest wyposażenie maszyn w moduły skutecznej komunikacji z użytkownikiem. Zadanie to realizuje się bardzo często z wykorzystaniem teorii zbiorów rozmytych i logik rozmytych (zob. np. (Terano, Asai, Sugeno 1994))³³.

Odpowiedź na pytanie drugie stanowi decyzję co do sposobu interpretacji teorii zbiorów rozmytych na polu semiotyki. Aby dać jakiś obraz rysujących się tu możliwości, rozważmy dwie sytuacje: (1) sytuację, w której zbiory rozmyte mają reprezentować przeciętne znaczenia terminów w pewnej społeczności językowej, oraz (2) sytuację, w której mają one odpowiadać znaczeniom przyjmowanym przez konkretne osoby.

W przypadku pierwszym – nazwijmy go „analizą języka publicznego” – dość naturalna wydaje się interpretacja statystyczna. Zgodnie z nią dla każdego terminu nieostrego należy ustalić, jaki procent członków danej społeczności językowej przypisuje mu konkretny obiekt (np. przedmiot materialny lub wielkość liczbową), a następnie zamienić ten procent na wartość funkcji przynależności. Weźmy za przykład termin *młody*. Przypuśćmy, że jego znaczenie będziemy modelować za pomocą zbioru rozmytego A ($A \in \mathcal{F}([10,60])$). Jeśli 10% ogółu uzna za młodą osobę w wieku 40 lat, to liczbie 40 przypiszemy $m_A(40) = 0,1$, jeśli 80% ogółu uzna za młodą osobę w wieku 30 lat, to przyjmujemy $m_A(30) = 0,8$ itd. Postępując tak, zrekonstruujemy przeciętny sposób rozumienia terminu *młody* w całej społeczności.

Kiedy jednak nasz wybór interpretacji odniesiemy nie do całej społeczności, ale jej konkretnego członka – nazwijmy to „analizą języka prywatnego” – będziemy musieli ustalić, z jakim przekonaniem (ewentualnie: z jaką częstością) ta właśnie osoba zalicza do młodych osoby w wieku od 10 do 60 lat. Oczywiście siła przekonania danej osoby nie musi być zgodna ze statystyką ogólną³⁴. W tym miejscu możemy przywołać ponownie operacjonistyczną interpretację pojęć, zgodnie z którą zdania typu „ X ma pojęcie M -a” powinny być rozumiane jako zdania czysto operacyjne w rodzaju „ X klasyfikuje obiekty jako M -y i nie- M -y”. W przypadku pojęć nieostrych decyzje klasyfikacyjne X -a odbiegają od schematu TAK-NIE, a dzieje się tak co najmniej z dwóch powodów: a) z powodu historii

³³ Niezależnie od wąskiego kontekstu badań nad systemami sztucznymi zasadnicza niemożność uściślenia języka naturalnego (czy też jakiegokolwiek języka odnoszącego się do rzeczywistości niewyidealizowanej) wynika z dwóch założeń ontologicznych, których przyjęcie wydaje się nieodparte: (1) język służący do opisu świata pozwala na utworzenie co najwyżej przeliczalnej nieskończoności wyrażeń, zaś liczba obiektów wyodrębnialnych w świecie ma moc continuum (krótko: język jest tworem dyskretnym, a świat – ciągłym), (2) pomiary wielkości ciągłych, zwłaszcza mikroskopowych, są zawsze obciążone błędami, co powoduje, że wyniki pomiarów (wyrażone w języku liczb) nigdy nie są ostre.

³⁴ W kontekście projektowania maszyn komunikujących się z ludźmi (w sposób możliwie naturalny) rozróżnienie między analizą języka publicznego i analizą języka prywatnego jest ważne z tego między innymi powodu, że projektując maszynę, należy zdecydować, czy ma być ona użytkowana przez konkretną osobą (i „dostosowana” do jej języka), czy też ma być użytkowana publicznie, przez różne osoby.

percepcyjnej X -a (to znaczy tego, jak X uczył się używać nazw i rozpoznawać przypisane do nich obiekty), b) z powodu kontekstu sytuacyjnego (który może wprowadzać elementy utrudniające jednoznaczną klasyfikację pewnych obiektów). W przypadku języka prywatnego większą rolę odgrywa czynnik pierwszy. To właśnie indywidualne obiekty, które X rozpoznawał trafnie jako desygnaty pojęcia (ale także jego porażki percepcyjne), ukształtowały jego indywidualny wzorzec percepcyjny, który z formalnego punktu widzenia możemy rozumieć jako funkcję przynależności zbioru rozmytego odpowiadającego danemu pojęciu.

Niezależnie od takiej czy innej interpretacji teorii zbiorów rozmytych największe korzyści, jakie może z nich odnieść teoria pojęć, mają charakter operacyjny. Chodzi o to, że omawiana teoria, a przede wszystkim postulowane przez nią operacje, umożliwia **precyzyjny rachunek pojęć nieostrych**. Mamy tu na myśli rachunek samych pojęć oraz rachunek zdań, w których pojęcia występują. W pierwszym przypadku teoria pozwala precyzyjnie stwierdzić, jak wyglądają rozmyte zakresy pojęć będących złożeniami pewnych pojęć podstawowych. Na przykład, jeśli założyć, że (1) zdefiniowano zbiory odpowiadające terminom *czerwony* i *pomarańczowy*, a ponadto (2) ustalono, że funktorom *i* oraz *nie* odpowiadają rozmyte operacje przecięcia i dopełnienia zbiorów, a funktorom *bardzo* i *raczej* – określone modyfikatory rozmyte, to teoria pozwala wyznaczyć zakresy takich pojęć, jak „*bardzo czerwony i nie pomarańczowy*” oraz „*raczej nie czerwony*”. W drugim przypadku – a więc wtedy, gdy wchodzi w grę nie same pojęcia, lecz wiążące je zdania – na pierwszy plan wysuwa się zagadnienie wnioskowania. Załóżmy dla przykładu, że stwierdzono fakt „Ten oto owoc jest bardzo zielony” oraz przyjęto regułę ogólną „Jeśli pewien owoc jest zielony, jest niedojrzały”. Na tej podstawie trzeba wywnioskować, jakiego rodzaju niedojrzałość charakteryzuje ten oto owoc, a więc, czy na przykład „jest on niedojrzały”, „raczej niedojrzały” czy może „bardzo niedojrzały”? Teoria zbiorów rozmytych daje nam do dyspozycji odpowiednie schematy wnioskowań, które są uogólnieniami schematów logik wielowartościowych, a polegają na składaniu relacji rozmytych reprezentujących zdania (w tym: implikacje). Schematy takie stosuje się szeroko w informatyce³⁵.

3.3. Pojęcia pewne i przybliżone

Omówiona w poprzednim podrozdziale teoria zbiorów rozmytych stanowi formalne narzędzie opisu pojęć nieostrych oraz metod przetwarzania informacji obejmujących takie pojęcia. Z ewentualnym zastosowaniem tej teorii w konkretnej dziedzinie wiąże się jednak pewien mankament. Otóż jakkolwiek pozwala ona

³⁵ Kwestii wnioskowania z danych rozmytych na podstawie rozmytej reguły odrywania poświęciłem odrębną pracę, (Stacewicz 1994). Por. też (Czogała, Pedrycz 1985).

manipulować pojęciami nieostryimi, to nie wyjaśnia, gdzie tkwi źródło ich nieostrości. Innymi słowy, funkcje przynależności odpowiednich zbiorów rozmytych są czymś, co się po prostu przyjmuje jako dane, a nie czymś, co trzeba dopiero skonstruować³⁶.

Zagadnienie pochodzenia nieostrości zostało postawione i częściowo rozwiązane w teorii zbiorów przybliżonych, której wyczerpujący wykład zawiera praca Zdzisława Pawlaka, „*Rough Sets. Theoretical Aspects of Reasoning about Data*” (Pawlak 1991)³⁷. W teorii tej przyjmuje się, iż obiektywne źródło nieostrości pojęć tkwi w naturalnej dysproporcji między zasobem wyrażen danego języka a liczbą możliwych do wyodrębnienia stanów świata (liczbą skończoną lub nieskończoną³⁸). Skupiając uwagę na nazwach, chodzi o sytuacje, w których **dany zestaw nazw o ostrych zakresach nie wystarcza do ścisłego opisu wszelkich dających się wyodrębnić grup obiektów**. Niektóre grupy, owszem, dają się opisać precyzyjnie – i te podpadają pod pojęcia ostre; inne natomiast są opisywalne tylko za pomocą przybliżeń – i te podpadają pod pojęcia rozmyte o możliwych do wyznaczenia rozmytych zakresach.

Prezentację podstaw teorii zbiorów przybliżonych zaczniemy od sugestywnego przykładu. Załóżmy, że dane jest uniwersum 15 przedmiotów, $U = \{x_1, \dots, x_{15}\}$, które scharakteryzowano dokładnie pod względem koloru. Przedmioty x_1, x_2 i x_3 są białe, przedmioty od x_4 do x_7 – żółte, przedmioty x_8, x_9 i x_{10} – pomarańczowe, a przedmioty od x_{11} do x_{15} – czerwone. Cecha *Kolor* (K) wyznacza zatem następujący podział uniwersum: $U/K = \{\{x_1, x_2, x_3\}, \{x_4, \dots, x_7\}, \{x_8, x_9, x_{10}\}, \{x_{11}, \dots, x_{15}\}\}$. Jego kolejne człony można traktować jako klasy obiektów nierozróżnialnych pod względem koloru, czyli klasy abstrakcji relacji równoważności indukowanej przez cechę K . Możemy też powiedzieć, że w danym uniwersum człony podziału są zakresami pojęć wyznaczonych przez cechę K ; w naszym przypadku pojęć *kolor biały*, *kolor żółty*, *kolor pomarańczowy*, *kolor czerwony*.

Założmy dalej, że ktoś wyodrębnił przedmioty od x_3 do x_9 jako żółtawe (a więc podobne do siebie ze względu na pewien odcień żółci) i zastanawia się, czy istnieje możliwość jednoznacznego opisu nowego pojęcia *żółtawy* za pomocą pojęć już określonych, tj. *biały*, *żółty*, *pomarańczowy* i *czerwony*? Możliwości takiej być nie może, ponieważ przedmioty wyodrębnione jako żółtawe, choć są z całą pewnością żółte i nie są z całą pewnością czerwone, to oprócz tego bywają

³⁶ W poprzednim podrozdziale sygnalizowaliśmy wprowadzić pewną metodę określania funkcji przynależności zbiorów rozmytych w kontekście języka publicznego – metodę odwołującą się do statystyki – niemniej sama teoria zbiorów rozmytych abstrahuje od tego rodzaju metod. Opisuje ona ogólne własności zbiorów oraz operacji na nich (pod warunkiem, że same zbiory są już określone).

³⁷ W poniższych uwagach będziemy odwoływać się niemal wyłącznie do tej pozycji; odnośniki do literatury będziemy umieszczać zatem tylko w tych miejscach, gdzie wykorzystamy informacje z innych źródeł.

³⁸ Jest to *de facto* kwestia szerszej dyskusji filozoficznej.

pomarańczowe i bywają białe. Mówiąc krótko: za pomocą pojęć już określonych pojęcie *żółtawy* daje się zdefiniować jedynie w przybliżeniu.

Ponieważ zbiór przedmiotów *żółtawych*, $A = \{x_3, \dots, x_9\}$, nie daje się przedstawić jako suma członów podziału U/K , nazywa się go *niedefiniowalnym* za pomocą cechy K lub krótko: *przybliżonym*. Przybliżają go dwie specyficzne sumy członów podziału U/K : przybliżenie dolne $A_D = \{x_4, \dots, x_7\}$ oraz przybliżenie górne $A_G = \{x_1, \dots, x_{10}\}$. Łatwo spostrzec, że do zbioru A_D należą przedmioty, które są „na pewno żółtawe”, zaś do A_G przedmioty, które „mogą być żółtawe”. W stylistyce „pojęciowej” powiedzielibyśmy, że pojęcie o zakresie A , czyli pojęcie żółtawości, jest pojęciem przybliżonym (w uniwersum U , ze względu na cechę *Kolor*)

Wyjaśnijmy jeszcze związek między przybliżonym charakterem zbioru A i jego nieostrością. Przyjmijmy w tym celu taką interpretację powyższego przykładu, w której osoba O_1 , wiedząca dokładnie co jest żółtawe, przybliża pojęcie *żółtawy* osobie O_2 , która nie zna wprawdzie pojęcia *żółtawy*, ale ma pewien zasób pojęć ostrych (tj. *biały*, *żółty*, *pomarańczowy* i *czerwony*). Gdy osoba O_2 uzyska od osoby O_1 informacje o przybliżeniu górnym i dolnym zbioru A (a nic więcej uzyskać nie może), będzie przejawiała następujące zachowania: a) będzie dobrze rozpoznawała jako żółtawe przedmioty żółte, b) będzie dobrze rozpoznawała jako nieżółtawe przedmioty czerwone, c) będzie myliła się (z punktu widzenia osoby O_1) co do żółtawości przedmiotów białych i pomarańczowych. Z perspektywy osoby O_2 pojęcie *żółtawy* jest więc nieostre, a zbiór A rozmyty. Okazuje się, co wyjaśnimy dalej, że jego dokładną funkcję przynależności daje się wyznaczyć na podstawie członów podziału X/K (które są zakresami pojęć pewnych).

Mając na uwadze powyższy przykład, możemy przedstawić koncepcję zbiorów przybliżonych w sposób ogólny, abstrahując od konkretnych uniwersów i cech (wyznaczających konkretne relacje równoważności).

Niech dana będzie *wiedza pewna* o danym uniwersum U (uniwersum przedmiotów, zjawisk, stanów itp.), czyli podział uniwersum U na niepuste i rozłączne klasy $X_1, \dots, X_n$ takie, że $X_i \subseteq U$, $X_i \cap X_j = \emptyset$ dla $i \neq j$, $\cup X_i = U$. Tak określoną wiedzę nazywa się również *klasyfikacją*, a klasy X_i jej *kategoriami*. Ponieważ poszczególne kategorie klasyfikacji są rozłączne i sumują się do dziedziny U , można je utożsamić z klasami abstrakcji pewnej relacji równoważności R . Z tego względu wiedzę oznacza się również jako U/R (co czytamy: U dzielone ze względu na R), gdzie $U/R = \{X_1, X_2, \dots, X_n\}$.

Zbiór X nazywa się *definiowalnym* w świetle wiedzy U/R , jeśli jest on sumą pewnych członów podziału U/R . W przeciwnym wypadku X jest *niedefiniowalny* lub *przybliżony*. Zbiory przybliżone charakteryzuje się poprzez ich przybliżenia dolne i górne, wyznaczone zawsze ze względu na wiedzę pewną U/R . *Przybliżenie dolne* X_D zbioru X tworzą wszystkie te elementy, które z całą pewnością można zaklasyfikować do X . Wyraża to warunek: $X_D = \cup \{Y \in U/R : Y \subseteq X\}$. *Przybliżenie górne* X_G zbioru X składa się z tych elementów dziedziny U , co do których nie

jest wykluczone, iż należą do X . Wyraża to warunek: $X_G = \cup(Y \in U/R: Y \cap X \neq \emptyset)$. Miarę dokładności zbioru X można określić jako $\alpha(X) = \text{card}(X_D)/\text{card}(X_G)$, gdzie card oznacza licznosc zbioru. Jeśli zbiór X uznamy za rozmyty (co wydaje się dość naturalne, zob. podrozdział 3.2), to miarę tę można wykorzystać prosto do określenia jego funkcji przynależności. Wszystkim elementom zbioru X_D wystarczy przypisać stopień przynależności 1, wszystkim elementom różnicy $X_D - X_G$ stopień przynależności $\alpha(X)$, zaś wszystkim pozostałym elementom uniwersum U wartość zerową. Por. (Pawlak, 1991).

Wróćmy jednak do pojęcia zbioru przybliżonego na gruncie wiedzy złożonej z pojęć pewnych. Gdy wiedza pewna W nie jest generowana przez jedną tylko relację równoważności R , ale przez całą rodzinę $\mathbf{R}$, przecięcie wszystkich relacji tworzących $\mathbf{R}$ nazywa się *relacją nierozróżnialności* (ang. *indiscernability*) i oznacza jako $\text{IND}(\mathbf{R})$. Klasy równoważności tej relacji, czyli klasyfikację $U/\text{IND}(\mathbf{R})$, otrzymuje się, wiedząc że $[x]_{\text{IND}(\mathbf{R})} = \bigcap_{R \in \mathbf{R}} [x]_R$. Termin „nierozróżnialność” oznacza, iż w kontekście wiedzy indukowanej przez rodzinę $\mathbf{R}$ elementy należące do poszczególnych kategorii relacji $\text{IND}(\mathbf{R})$ pozostają nierozróżnialne. Przykładowo, jeśli rodzina $\mathbf{R}$ składa się z relacji odpowiadających trzem cechom przedmiotów – kolorowi, kształtowi i wielkości – to klasy abstrakcji relacji $\text{IND}(\mathbf{R})$ tworzą zbiory obiektów nierozróżnialnych ze względu na te cechy, mające zatem ten sam kolor, kształt i wielkość.

W wiedzy indukowanej przez pewną rodzinę relacji $\mathbf{R}$, czyli $U/\mathbf{R}$, wyróżnia się wiedzę *podstawową* i *elementarną*. Na pierwszą składają się kategorie podziałów indukowanych przez poszczególne relacje rodziny $\mathbf{R}$, na drugą – kategorie podziału $U/\text{IND}(\mathbf{R})$. Jeśli w rodzinie $\mathbf{R}$ zostanie wyróżniona pewna podrodzina $\mathbf{P}$, to kategorie podziału $U/\text{IND}(\mathbf{P})$ nazywa się wiedzą $\mathbf{P}$ -elementarną.

Wyróżnienie w rodzinie $\mathbf{R}$ relacji (cech) warunkowych $\mathbf{P}$ i decyzyjnych $\mathbf{Q}$ ($\mathbf{P} \cup \mathbf{Q} = \mathbf{R}$) oznacza takie przeformułowanie problemu wiedzy, w ramach którego można opisywać wnioskowania. Prowadzą one od stwierdzenia konkretnych wartości cech warunkowych obiektu do decyzji o przysługiwaniu mu odpowiednich wartości cech decyzyjnych. Wiedza jest rozumiana wówczas jako możliwość wnioskowania o przynależności obiektów do kategorii podziału $U/\text{IND}(\mathbf{Q})$ na podstawie ich przynależności do kategorii podziału $U/\text{IND}(\mathbf{P})$. Powiązanie między kategoriami tych dwóch podziałów można wykorzystywać w przypadku nowych obiektów: za każdym razem kiedy okaże się, iż obiekt należy do pewnej kategorii podziału $U/\text{IND}(\mathbf{P})$, czyli ma pewne wartości cech warunkowych, musi należeć do odpowiedniej kategorii podziału $U/\text{IND}(\mathbf{Q})$, czyli mieć pewne wartości cech decyzyjnych.

W przypadku wiedzy indukowanej nie przez pojedynczą cechę/relację, lecz przez całą ich rodzinę, definiowalność zbiorów (zakresów pojęć) określa się w odniesieniu do zestawu kategorii elementarnych, czyli podziału $U/\text{IND}(\mathbf{R})$. Zbiór X nazywa się pewnym lub definiowalnym, jeśli jest on sumą kategorii elementarnych. Definicje przybliżeń dolnych i górnych modyfikuje się adekwatnie

do zastąpienia podziału U/R podziałem $U/IND(\mathbf{R})$. Zatem: przybliżenie dolne zbioru X to $X_D = \cup(Y \in U/IND(\mathbf{R}): Y \subseteq X)$, zaś przybliżenie górne zbioru X to $X_G = \cup(Y \in U/IND(\mathbf{R}): Y \cap X \neq \emptyset)$.

W bezpośrednim związku z przedstawioną powyżej teorią pozostają konkretne metody reprezentacji wiedzy, czyli określonego układu pojęć podstawowych i ostrych zarazem (Słowiński 1992). Do metod najczęściej wykorzystywanych, zwłaszcza w kontekście systemów informatycznych, należą tabele wartości cech (w tym: tabele decyzyjne) oraz formuły klasycznego rachunku zdań uzupełnionego o pewną interpretację formuł atomowych (rachunek taki nazywa się niekiedy logiką decyzyjną, zob. (Pawlak, 1991; Piętka, Stacewicz 2015). Podstawą obydwu sposobów reprezentacji są konkretne cechy opisywanych obiektów i ich wartości (odpowiadają one relacjom z rodziny $\mathbf{R}$, o której była mowa powyżej). Każda z cech wyznacza klasyfikację, której kategoriami są zbiory obiektów mających poszczególne wartości cechy. Na przykład, jeśli weźmiemy pod uwagę cechę *Kolor* o wartościach *Żółty*, *Czerwony* i *Biały*, to pierwsza z wyznaczanych przezeń kategorii zawiera wszystkie obiekty żółte, druga – wszystkie czerwone, a trzecia – białe.

Przedmiot	Kształt (Ks)	Wielkość (W)	Kolor (Kl)
x_1	<i>okrągły</i>	<i>średni</i>	<i>czerwony</i>
x_2	<i>okrągły</i>	<i>mały</i>	<i>czerwony</i>
x_3	<i>okrągły</i>	<i>mały</i>	<i>biały</i>
x_4	<i>podłużny</i>	<i>duży</i>	<i>biały</i>
x_5	<i>podłużny</i>	<i>duży</i>	<i>żółty</i>
x_6	<i>kwadratowy</i>	<i>średni</i>	<i>biały</i>
x_7	<i>kwadratowy</i>	<i>duży</i>	<i>żółty</i>
x_8	<i>kwadratowy</i>	<i>duży</i>	<i>żółty</i>

Metodę tabelaryczną zilustrujemy prostym przykładem zapisu wiedzy dotyczącej ośmiu przedmiotów charakteryzowanych przez trzy cechy.

Kolejne cechy generują następujące podziały:

$$U/Ks = \{\{x_1, x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7, x_8\}\}$$

$$U/W = \{\{x_4, x_5, x_7, x_8\}, \{x_2, x_3\}, \{x_1, x_6\}\}$$

$$U/Kl = \{\{x_1, x_2\}, \{x_3, x_4, x_6\}, \{x_5, x_7, x_8\}\}$$

Wyszczególnionym podziałom odpowiada relacja nierozróżnialności $IND(Ks, W, Kl)$, której kolejne kategorie (elementarne) pokrywają się z zakresami pojęć pewnych – pewnych ze względu na wiedzę indukowaną przez wyróżnione powyżej cechy. Oto stosowny podział:

$$U/IND(Ks, W, Kl) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7, x_8\}\}$$

Inna relacja nierozróżnialności, nierozróżnialności ze względu na cechy *Kształt* i *Wielkość*, czyli $U/IND(Ks, W)$, określa mniej szczegółowy podział uniwersum niż powyżej:

$$U/IND(Ks, W) = \{\{x_1\}, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6\}, \{x_7, x_8\}\}$$

Na potrzeby formalizacji wnioskowań (a także ich implementacji) zapis tabelaryczny wiedzy wygodnie jest przekształcić w zapis logiczny (zob. (Pawlak, 1991; Piętko, Stacewicz 2015)). Zgodnie z oryginalną koncepcją Z. Pawlaka, na zapis ten składają się formuły atomowe typu (*cecha, wartość*) oraz formuły złożone, które powstają przez łączenie formuł atomowych za pomocą spójników klasycznego rachunku zdań. Przynajmniej niektóre z formuł (w tym: atomowe) można interpretować jako nazwy, których treścią są pewne pojęcia, zaś zakresami zbiory obiektów uniwersum U . Oto przykłady kilku formuł odnoszących do prezentowanej powyżej tabeli:

(*kształt, podłużny*), (*kolor, biały*) – formuły atomowe

(*kształt, okrągły*) $\vee$ (*wielkość, mały*) – formuła złożona (alternatywa)

(*kształt, okrągły*) $\wedge$ (*wielkość, mały*) $\wedge$ (*kolor, biały*) – formuła złożona (konjunkcja odpowiadająca trzeciemu wierszowi tabeli).

Każda z formuł, interpretowana jako nazwa prosta lub złożona, ma określony zbiór desygnatów, czyli denotację. Denotacje nazw prostych są kategoriami podstawowymi podziałów indukowanych przez występujące w tych nazwach cechy. Na przykład, denotacją formuły (*kształt, okrągły*) jest zbiór $\{x_1, x_2, x_3\}$. Denotacje wyrażeń złożonych wyznacza się, wykonując na kategoriach podstawowych działania teoriomnogościowe odpowiadające spójnikom logicznym. Na przykład, zbiór desygnatów formuły $\varphi := [(*kształt, podłużny*) \wedge (*kolor, biały*)] \vee (*wielkość, mała*)$ wyznacza się, wykonując działania $(\{x_4, x_5\} \cap \{x_3, x_4, x_6, x_7\}) \cup \{x_2, x_3\}$ i otrzymując $\{x_2, x_3, x_4\}$.

Wskazana metoda określania denotacji formuł dotyczy wiedzy pewnej, odnoszącej do dobrze określonych podziałów uniwersum za pomocą takich a nie innych cech (w powyższym przykładzie kształtu, wielkości i koloru). Nic nie stoi jednak na przeszkodzie, aby do języka systemu wprowadzić terminy nieostre (jak żółtawy), których znaczenia odpowiadają pojęciom przybliżonym (niedefiniowalnym w ramach zastanej wiedzy W). Jak wyjaśnialiśmy wcześniej, zakresy takich nazw są zbiorami rozmytymi. Z tego powodu, wyznaczając zakresy takich nazw złożonych, w których występują terminy nieostre, należy wykonywać operacje na zbiorach rozmytych. Ich wyniki są wprawdzie zbiorami rozmytymi, ale zgodnie

z rozważaną tu teorią, dla każdego z takich zbiorów można podać dwa przybliżenia ostre, przybliżenie dolne i górne. To zaś daje nam więcej informacji o pojęciu niż sam fakt stwierdzenia jego nieostrości.

Zapis wiedzy w postaci formuł logicznych jest szczególnie użyteczny wtedy, kiedy wśród cech wyróżnia się cechy warunkowe i decyzyjne. Wówczas zależności między wartościami cech przedstawia się w postaci implikacji np. (*kształt, okrągły*) $\wedge$ (*wielkość, średni*) $\rightarrow$ (*kolor, czerwony*). Zbiory takich implikacji stanowią podstawę wnioskowań o przysługiwaniu obiektom określonych cech decyzyjnych (po uprzednim stwierdzeniu określonych wartości cech warunkowych). Ciekawą własnością opisywanego systemu logicznego jest możliwość algorytmicznej redukcji pierwotnego zbioru implikacji, tak aby uzyskać zbiór istotnie mniejszy (zawierający mniejszą liczbę implikacji i/lub mniej złożone implikacje), lecz zapewniający taką samą skuteczność wnioskowań (Bolc, Cytowski Stacewicz 1996).

Reasumując to, co przedstawiono powyżej, możemy stwierdzić, że teoria zbiorów przybliżonych wraz z towarzyszącą jej metodą tabelarycznej reprezentacji danych oraz logiką decyzyjną, umożliwiają precyzyjny rachunek pojęć przybliżonych (interpretowalnych niekiedy jako pojęcia nieostre). Pozwalają także osadzić reprezentowane formalnie pojęcia w pewnym rachunku logicznym, który zapewnia efektywne schematy wnioskowań na podstawie wyrażanej za ich pomocą wiedzy.

3.4. Pojęcia zbiorowe i niezbiorowe

W dwóch poprzednich podrozdziałach termin *zbiór* był używany w znaczeniu podstawowym i najbardziej rozpowszechnionym: znaczeniu *dystrybutywnym*. Zgodnie z nim za zbiór uznaje się grupę elementów wyróżnionych z ogółu i zestawionych razem ze względu na przysługującą im wspólnie własność. Na przykład, wyrazy wyodrębnione z poprzedniego zdania ze względu na „dwuliterowość” tworzą zbiór czteroelementowy {„za”, „ze”, „na”, „im”}³⁹.

Podobnie jak własności wyznaczające zbiory są czymś ontologicznie pierwotnym, tak same zbiory nie dadzą się objaśnić za pomocą pojęć prostszych. Fakt ten ujawnia się wyraźnie na polu matematyki, gdzie pojęcie zbioru poddaje się najdalej idącej precyzacji. Jak wiadomo, odpowiednia definicja ma charakter

³⁹ Gwoli ścisłości trzeba zauważyć, że odwołanie się do wspólnej cechy elementów (strategia intensionalna) nie stanowi jedynej metody określania zbiorów. Niekiedy określa się je po prostu przez wskazanie pewnych elementów, bez wyjaśniania dlaczego, czyli ze względu na jaką własność, zostały one zebrane razem (strategia ekstensionalna). W tym drugim przypadku albo w ogóle nie można twierdzić, iż zbiory odpowiadają własnościom przedmiotów, albo można tak twierdzić, dodając przy tym, że na etapie określania zbiorów własności te nie są znane.

aksjomatyczny, a tym samym nie odwołuje się do terminów już zdefiniowanych, lecz polega na uznaniu za prawdziwe pewnych zdań, w których obok terminu *zbiór* figurują wyłącznie inne terminy pierwotne, takie jak *element* i *należy do*. Wiadomo nadto, że aksjomaty precyzujące pojęcie zbioru stanowią podstawę teorii mnogości; na jej fundamencie zaś wspiera się cały gmach współczesnej matematyki (Kuratowski, Mostowski 1986).

Ujmując rzecz psychologicznie, do zbiorów w sensie dystrybutywnym prowadzi proces *abstrakcji*. Abstrahując od pewnych własności, czyli pomijając je myślowo, a wybijając na pierwszy plan inne, człowiek skupia uwagę na bytach innego rodzaju niż same obiekty (tj. zbiorach), a następnie byty te czyni podstawą dalszych czynności poznawczych. Kwestią sporną, filozoficzną już, pozostaje natomiast, czy wynik procesu abstrakcji istnieje w takim samym sensie co materiał, na którym abstrakcja się dokonuje. W szczególności, zagadnienie to wyraża się następującym pytaniem: „Jeśli przyjmiemy zgodnie ze zdrowym rozsądkiem, że pewne obiekty istnieją realnie, czyli niezależnie od konstrukcyjnej działalności umysłu, to czy mamy prawo przyjąć, że równie realnie istnieją zbiory tychże obiektów – nie oddzielne obiekty, ale nowe byty w postaci zbiorów?”.

Na podłożu wątpliwości, co do pozytywnej odpowiedzi na przytoczone powyżej pytanie, rozwinęło się drugie, nieco mniej typowe, rozumienie zbiorów. Polega ono na uznawaniu za zbiory nie pewnych bytów odpowiadających wyabstrahowanym przez umysł własnościom przedmiotów, ale pewnych realnych *całości* złożonych z części. Byty tego rodzaju nazywa się zbiorami w sensie *kolektywnym* lub *mereologicznym*, a do ich typowych przykładów zaliczają się: księgozbiór złożony z książek, łańcuch złożony z ogniw, rój złożony z pszczoł.

Logikiem, który zapewnił nowemu pojęciu zbioru solidne podstawy matematyczne – uważane przez niektórych za alternatywę dla tradycyjnej teorii mnogości – był Stanisław Leśniewski⁴⁰. Na podstawy te składają się trzy systemy dedukcyjne – prototypyka, ontologia i mereologia – spośród których ostatni, zależny logicznie od dwóch pierwszych, stanowi właściwą formalną teorię części i całości. Sformułowana w ramach mereologii definicja zbioru (mnogości) brzmi następująco: „Wyrażenia „*mnogość przedmiotów m*” używam dla oznaczenia każdego takiego przedmiotu *P*, który czyni zadość następującemu warunkowi: jeżeli *I* jest ingrediensem przedmiotu *P*, to pewien ingrediens przedmiotu *I* jest ingrediensem pewnego *m*, które jest ingrediensem przedmiotu *P*” (Leśniewski 1916). Definicja ta staje się zrozumiała w pełni dopiero na gruncie dwóch pierwszych aksjomatów mereologii,

⁴⁰ Matematyczne podstawy ogólnej teorii mnogości Stanisława Leśniewskiego (zwanej mereologią), wraz z towarzyszącymi im uwagami natury metodologiczno-filozoficznej, zostały przedstawione w latach 1927–1931, w cyklu publikacji pod wspólnym tytułem „O podstawach matematyki”, w Przeglądzie Filozoficznym (roczniki od XXX do XXXIV). Publikację tę poprzedziła wczesna praca z roku 1916, p.t. „Podstawy ogólnej teorii mnogości. I.”. Szczegółowe informacje o tych publikacjach znajdują się w bibliografii.

które precyzują formalne własności relacji „bycia częścią”, a więc w szczególności „bycia ingrediensem”. Co ważne jednak, określenie każdego konkretnego zbioru (w sensie kolektywnym) wymaga, po pierwsze, wskazania, z jakiego rodzaju przedmiotów się on składa, i, po drugie, określenia konkretnej relacji bycia częścią. Przykładowo: określając znaczenie nazwy „łańcuch”, należy stwierdzić, iż jest to całość złożona z tak a nie inaczej łączonych ogniw. Por. (Skowron, 2021).

Do objaśnionej powyżej różnicy między zbiorami dystrybutywnymi (po prostu: zbiorami) a zbiorami kolektywnymi (całościami) odwołuje się popularna typologia nazw, w myśl której rozróżnia się nazwy *zbiorowe* i *niezbiorowe*. Różnica między nimi ujawnia się na poziomie desygnatów. O ile bowiem desygnatami pojęć *niezbiorowych* są przedmioty indywidualne, tj. wyróżniane bez względu na wiążące je relacje i ich wewnętrzną strukturę, o tyle desygnatami pojęć *zbiorowych* są całości o pewnej strukturze wewnętrznej, czyli układy przedmiotów powiązanych ze sobą takimi relacjami, iż narzuca się ujęcie ich w całość (zob. np. Ziemiński 2000). Na przykład, ta oto szachownica, będąca typowym desygnatem pojęcia *zbiorowego*, to układ białych i czarnych pól ułożonych na przemian, ściśle przylegających do siebie, a tworzących razem kwadrat o takich a takich rozmiarach. O specyficznym statusie desygnatów nazw *zbiorowych* informuje treść: jej istotny składnik stanowi bowiem cecha bycia całością złożoną z takich a takich części. Na przykład nazwa „łańcuch” oznacza różne całości złożone w ściśle określony sposób z ogniw, a do jej treści oprócz cech dotyczących całego łańcucha (np. *kształt podłużny*) należy również cecha *bycia całością złożoną z ogniw*.

Mając na względzie powyższe uwagi, możemy powiedzieć, co następuje. Zakresami wszelkich pojęć (*zbiorowych* i *niezbiorowych*) są zbiory dystrybutywne. Desygnatami pojęć *zbiorowych* są całości złożone z części, czyli pewne zbiorowości, zwane zbiorami kolektywnymi. W konsekwencji teoria zbiorów kolektywnych jest niezbędna tylko na etapie definiowania funkcji decyzyjnej odpowiadającej pojęciu *zbiorowemu* – funkcji w pewnym sensie odpowiedzialnej za rozpoznawanie desygnatów będących całościami (nie jest natomiast przydatna przy analizie zakresów tychże pojęć). Mówiąc ogólnie, teoria ta daje wskazówki co do tego, jak odróżniać desygnaty pojęcia *zbiorowego* (będące takimi a takimi całościami) od innych obiektów. Mówiąc konkretniej, w formule funkcji decyzyjnej odpowiadającej pojęciu *zbiorowemu* musi być „zaszyta” informacja zarówno o konkretnej relacji bycia częścią, jak i o rodzaju przedmiotów tworzących całość.

Teoria zbiorów kolektywnych byłaby więc wykorzystywana pośrednio i tylko w przypadku pojęć *zbiorowych*. Do analizy zakresowej, a więc analizy tego, jak mają wyglądać zakresy pojęć (np. ostre czy nieostre) oraz operacje na nich, wystarczy, naszym zdaniem, stosować różne warianty teorii zbiorów dystrybutywnych.

3.5. Podsumowanie

1. Teorie pojęć rozumianych jako funkcje decyzyjne znajdują swoje teoretyczne podstawy w różnych **teoriach zbiorów**, niekoniecznie w teorii klasycznej. Zakres pojęcia jest zbiorem pewnego typu, zaś samo pojęcie uznaje się za funkcję charakterystyczną tego zbioru.
2. Klasyczna **teoria zbiorów** (ostrych) i powiązana z nią klasyczna logika dwuwartościowa służą do opisu pojęć ostrych, które są funkcjami typu $f: X \rightarrow \{0,1\}$, gdzie X jest uniwersum desygatów pojęć.
3. Teoria **zbiorów rozmytych** – występująca w różnych wariantach, opartych na różnych logikach wielowartościowych – służy do opisu pojęć nieostrych. Zapewnia rachunek takich pojęć oraz schematy wnioskowań dla zdań zawierających terminy nieostre.
4. Teoria **zbiorów przybliżonych** umożliwia definiowanie pojęć na podstawie wiedzy pewnej (niewątpliwej); zazwyczaj w taki sposób, że zakresy pojęć są określone przez ich przybliżenia dolne i górne (są to wówczas zakresy nieostre). Skojarzona z tą teorią **logika decyzyjna** umożliwia redukcję wiedzy do postaci zapewniającej taką samą efektywność wnioskowania jak wiedza pierwotna.
5. Teoria **zbiorów kolektywnych**, czyli mereologia Leśniewskiego, pozwala opisywać funkcje decyzyjne odpowiadające pojęciom zbiorowym. Każdy desygnat pojęcia zbiorowego jest całością złożoną z części, czyli zbiorem kolektywnym.

Rozdział 4

Pojęcia w perspektywie informatycznej

Informatyczny sposób opisu pojęć jest nieodłącznie związany z własnościami systemów sztucznych, które spełniają funkcje podobne do ludzkiego umysłu, na przykład: rozpoznają obiekty, tworzą ich klasyfikacje czy realizują wnioskowania skutkujące nową wiedzą. Ponieważ systemy takie działają na podstawie pewnych algorytmów, reprezentowana wewnątrz nich wiedza pojęciowa (niezależnie od tego czy przypomina wiedzę ludzką, czy jest czymś jakościowo odrębnym), musi być zapisana formalnie, za pomocą precyzyjnych struktur danych. To sprawia, że *pojęcia rozumiane informatycznie* przypominają pojęcia w sensie logicznym. Jednocześnie, ze względu na fakt, że wiele systemów informatycznych służy do modelowania ludzkich czynności poznawczych, zakodowane w tych systemach struktury pojęciowe można interpretować psychologicznie, jako struktury odpowiadające pojęciom w sensie psychologicznym. Sam fakt definiowania takich struktur, a więc posługiwania się konkretnymi formalnymi reprezentacjami pojęć, przesądza o tym, że podejście informatyczne ma charakter intensjonalny (a nie ekstensjonalny). Pojęcia są tu wprawdzie rozumiane jako funkcje przyporządkowujące obiektom z pewnej dziedziny decyzje klasyfikacyjne (co jest zgodne z tradycją Fregego), lecz funkcje te mają zawsze konkretny kształt. Ich formuły nie są czymś abstrakcyjnym, lecz programistycznie określonym.

4.1. Obszary badawcze i zagadnienia

W rozdziale pierwszym uzasadnialiśmy tezę, że pojęcia stanowią podstawowy składnik ludzkiej wiedzy – zarówno w ujęciu faktualnym, czyli psychologicznym, jak i normatywnym, czyli logicznym. Znaczy to między innymi tyle, że leżą one u podstaw bardzo wielu procesów poznawczych. Na pierwszy plan wysuwa się tutaj *percepcja*, czyli rozpoznawanie obiektów i zaliczanie ich do właściwych kategorii. Z psychologicznego punktu widzenia proces ten polega na uaktywnieniu odpowiedniej *reprezentacji pojęciowej*, która w sposób uogólniony reprezentuje w pamięci podmiotu pewien zbiór obiektów. W warstwie językowej czynności tej odpowiada skojarzenie z rozpoznawanym przedmiotem nazwy, której znaczeniem

jest właśnie pojęcie. Gdy weźmiemy pod uwagę różne typy wnioskowań, to pojęcia są dla nich podstawowe o tyle, że zapewniają określony sposób rozumienia przesłanek. Podobnie jest w przypadku uczenia się: nabycie odpowiednich pojęć daje możliwość przyswajania bardziej zaawansowanych treści.

Tak silny związek pojęć z różnymi formami zdobywania wiedzy sprawia, że informatyczne badania nad nimi **sytuują się głównie w obszarze sztucznej inteligencji (SI)**, czyli w tym dziale informatyki, który ma na celu algorytmizację i automatyzację ludzkich czynności poznawczych (począwszy od percepcji a skończywszy na wyrafinowanych rozumowaniach matematycznych; por. (Russel, Norvig 2020)). Jeśli za trzy główne zadania badawcze w dziedzinie SI uznać: a) *podejmowanie decyzji*, b) *automatyzację uczenia się*, c) *efektywną komunikację z człowiekiem*, to badania nad pojęciami są niezbędne do osiągnięcia każdego z tych celów (Stacewicz 2010). Dwa pierwsze zadania są ze sobą sprzężone i trudno sobie wyobrazić ich realizację bez zakodowania w systemie pojęć, które stanowią podstawowy składnik wiedzy systemu. Zadanie trzecie obejmuje rozumienie lub przynajmniej właściwe reagowanie na wypowiedzi użytkownika, a te muszą zawierać nazwy odnoszące do określonych grup przedmiotów (ich znaczenia to pojęcia nominalne). Ze względu na specyfikę komunikacji w języku naturalnym, obejmującym terminy nieostre, w rozwiązaniach informatycznych stosuje się aparat teorii zbiorów rozmytych i przybliżonych (zob. podrozdziały 3.3 i 3.4). Niezbędną podstawą komunikacji jest percepcja, czyli rozpoznawanie obiektów, w tym obrazów (szerzej: scen) i dźwięków (szerzej: słów). Tę zaś umożliwiają wbudowane w system lub nabyte w drodze uczenia się reprezentacje pojęciowe.

Zakodowana w systemach SI wiedza pojęciowa ma określoną postać, wynikającą z przyjętej w danym systemie ogólnej *metody reprezentacji wiedzy* (np. regułowej lub koneksyjnej). Na wyższym poziomie ogólności, abstrahującym od metody reprezentacji, **pojęcia są definiowane w duchu klasycznego ujęcia Fregego jako funkcje decyzyjne** (zob. rozdz. 2)⁴¹. Funkcje te przyporządkowują obiektom z pewnej dziedziny jednoznaczne lub stopniowalne decyzje kategoryzacyjne: „TAK, obiekt należy do zakresu pojęcia”, „NIE, obiekt nie należy do zakresu pojęcia”, ewentualnie „stopniowalne TAK/NIE, obiekt należy do zakresu pojęcia w pewnym stopniu”. W większości zastosowań są to funkcje binarne typu $p: U \rightarrow \{0,1\}$ ⁴² – gdzie U oznacza dziedzinę desygnatów pojęć, a $\{0,1\}$ jest

⁴¹ Ujęcia wywodzące się z tradycji fregowskiej nie są jedynymi obecnymi w zastosowaniach informatycznych. Oprócz nich rozważa się, na przykład, koncepcję pojęć ekstensjonalno-intensjonalnych w postaci par (obiekt, zbiór atrybutów), która została zapoczątkowana przez Rudolpha Willego, a jest obecnie rozwijana pod nazwą „Formal Concept Analysis”. Por. (Ganter, Wille 1999).

⁴² Zastosowania mniej typowe dotyczą wiedzy prawdopodobnej, której podstawowy składnik stanowią pojęcia nieostre, o zakresach będących zbiorami rozmytymi (zob. podrozdział 3.3). W przypadku tego rodzaju zastosowań systemy komputerowe generują roboczo szereg decyzji klasyfikacyjnych o różnych stopniach pewności (odpowiada im zbiór rozmyty na skończonym uniwersum), z których finalnie jest wybierana decyzja o najwyższym stopniu pewności (z zakresu rozmytego wybiera się jeden ostry element). Por. (Czogała, Pedrycz 1985).

dwuelementowym zbiorem decyzji klasyfikacyjnych. Przy takim zapisie zbiór $Z = \{x: p(x) = 1\}$ stanowi zakres pojęcia p (czyli zbiór jego desygnatów), a konkretna formuła $p(x)$ to jego treść, tożsama *de facto* ze zbiorem instrukcji/struktur odpowiedniego programu komputerowego.

Ponieważ zwykle obiekty rozpoznawane lub klasyfikowane przez system komputerowy są opisywane w kategoriach ich wyróżnionych cech, funkcja p stanowi przeważnie superpozycję funkcji opisu $c: X \rightarrow C_1 \times C_2 \times \dots \times C_n$ (gdzie C_i jest zbiorem wartości i -tej cechy) oraz funkcji decyzyjnej $d: C_1 \times C_2 \times \dots \times C_n \rightarrow [0,1]$ (określonej na iloczynie kartezjańskim zbiorów wartości cech)⁴³. Jej zadany przebieg, czy mówiąc inaczej, zadany rozkład przyporządkowań obiektom liczb 0 i 1, nie przesądza oczywiście o jej postaci, czyli kształcie formuły wiążącej wartości cech $C_1, \dots, C_n$ (traktowane jako zmienne). Formuł takich może być wiele, czemu w konkretnych przypadkach odpowiadają różne zapisy wewnątrz programów komputerowych. Informatyczny odpowiednik formuły d , czyli schematyczny sposób zakodowania jej za pomocą określonej metody reprezentacji wiedzy, nazywa się *informatyczną reprezentacją pojęcia*.

Przyjęte powyżej konwencje terminologiczne pozwalają sformułować **cztery ogólne zagadnienia dotyczące pojęć, które odpowiadają zadaniom operujących pojęciami systemów informatycznych**. Chociaż systemy te wykorzystują zawsze konkretne metody reprezentacji wiedzy i działają na podstawie konkretnych algorytmów, to same zagadnienia można opisać ogólnie. Por. (Cichosz 2000; s. 74-87).

1. Rozpoznawanie desygnatów pojęcia. Jest to zagadnienie czysto operacyjne, będące informatycznym odpowiednikiem psychologicznego problemu kategoryzacji (a szerzej: percepcji). Zakładając, że pojęcie p ma określoną reprezentację obliczeniową, od programu oczekuje się obliczania wartości $p(x)$ dla różnych obiektów x (a konkretniej: dla różnych wartości cech opisujących obiekty x). Wartość $p(x) = 1$ oznacza obiekt rozpoznany jako desygnat pojęcia, wartość $p(x) = 0$ – fakt przeciwny. W przypadku systemów operujących na informacji niepewnej dopuszcza się inne wartości $p(x)$, określające stopień pewności co do decyzji o rozpoznaniu desygnatu pojęcia (najczęściej: $p(x) \in [0,1]$)

2. Uczenie się pojęcia na podstawie przykładów. Pojęcie p nie ma jeszcze określonej reprezentacji obliczeniowej, ale są dane przykłady pozytywne jego desygnatów (zebrane w zbiorze $P = \{x: p(x) = 1\}$) oraz przykłady negatywne (zebrane w zbiorze $Q = \{x: p(x) = 0\}$). Od programu uczącego się oczekuje się wygenerowania takiej hipotetycznej reprezentacji pojęcia p , która zapewniłaby poprawne

⁴³ W przypadku pojęć ostrych przeciwdziedzina funkcji c jest zbiór $\{0,1\}$. W przypadku pojęć nieostrych (wiedzy niepewnej) zbiór ten jest rozszerzany do zbioru obejmującego więcej elementów należących do przedziału $[0,1]$, czasem nawet do całego przedziału $[0,1]$.

rozpoznawanie obiektów z P i Q (dla każdego $x \in P$ program ma zwracać wartość 1, dla każdego $x \in Q$ – wartość 0). Niekiedy jest określone pewne kryterium poprawnego rozpoznawania obiektów spoza zbiorów P i Q . Zagadnienie to stawia się często w wersji dynamicznej, to znaczy żąda się modyfikowania reprezentacji pojęciowej wraz z napływem nowych przykładów pozytywnych i negatywnych. W przypadku systemów operujących na informacji niepewnej do poszczególnych przykładów dołącza się ich stopnie pewności (co do przynależności obiektu do zakresu pojęcia), zaś zbiór wartości generowanego pojęcia-funkcji rozszerza się poza zbiór $\{0,1\}$ (zob. punkt 1).

3. Optymalizacja reprezentacji pojęcia. Zagadnienie polega na modyfikowaniu wygenerowanej wstępnie, np. w drodze uczenia się na podstawie przykładów, reprezentacji danego pojęcia. Programowi optymalizującemu stawia się zadanie wygenerowania takiej reprezentacji, aby zostało spełnione pewne kryterium (np. minimalna liczba wykorzystywanych cech i ich wartości). Zagadnienie to rozwiązuje się przy określonej dla danego systemu obliczeniowego metodzie reprezentacji wiedzy (np. metodzie regułowej lub konekcyjnej), a wspomniane powyżej kryterium dotyczy własności tej reprezentacji (np. liczby reguł).

4. Tworzenie pojęcia (przy braku wskazanych *a priori* przykładów pozytywnych i negatywnych). Jest to zagadnienie najszerze. W pierwszej kolejności stawia się wymóg podziału dziedziny X na dwa zbiory, P oraz $X-P$; pierwszy z nich ma zawierać desygnaty tworzonego pojęcia (przykłady pozytywne), drugi pozostałe obiekty dziedziny (przykłady negatywne). Zadaje się przy tym pewne kryterium podziału, najczęściej jest to kryterium maksymalizacji podobieństwa elementów z P oraz minimalizacji podobieństwa między obiektami z P oraz $X-P$ (relacja podobieństwa jest dana). W drugiej kolejności od programu tworzącego pojęcie wymaga się wygenerowania, na podstawie pozytywnych i negatywnych przykładów, optymalnej reprezentacji pojęcia. Wymaga to rozwiązania zagadnień 2 i 3.

Należy podkreślić, że cel określony w zagadnieniu pierwszym, czyli poprawne rozpoznawanie desygnatów pojęcia, stanowi jednocześnie kryterium poprawności dla rozwiązań pozostałych zagadnień. Zarówno pojęcia tworzone od podstaw (przy uprzednim podziale dziedziny X), jak i pojęcia generowane na podstawie przykładów, jak również i pojęcia optymalizowane ze względu na pewne własności ich reprezentacji, muszą zapewniać poprawne rozpoznawanie obiektów (w tym: pozytywnych przykładów ze zbioru P).

Ponadto, w przypadku realnych zastosowań tak czy inaczej reprezentowane pojęcia służą przede wszystkim poprawnej i efektywnej kategoryzacji w określonym uniwersum obiektów. W kontekście badań nad sztuczną inteligencją moglibyśmy powiedzieć, że **podstawową czynnością poznawczą, której automatyzacji**

służy formalizacja pojęć, jest percepcja. Inne czynności angażujące pojęcia, takie jak rozumowanie czy uczenie się, zawsze w jakiś sposób są powiązane z percepcją (np. celem uczenia się są poprawne decyzje kategoryzacyjne). Ich algorytmizacja⁴⁴ zaś, choć wymaga dodatkowych elementów/procedur, musi uwzględniać pewne schematy rozpoznawania obiektów.

W związku z automatyzowanymi przez informatyków procesami posługiwania się pojęciami – które ostatecznie prowadzą do pewnych decyzji kategoryzacyjnych – do zestawu zagadnień 1-4 należy dołączyć zagadnienie piąte, szczególnie ważne z punktu widzenia dalszej części książki. Jest to *zagadnienie wyjaśniania*. Sytuuje się ono na przecięciu informatyki, psychologii i epistemologii, a wynika z konieczności opracowania specjalnych procedur, które skutkowałyby przekonującymi dla człowieka wyjaśnieniami podejmowanych decyzji. Przykładowo: ekspercki system medyczny, który rozpoznaje występujące u pacjenta objawy jako wskazujące na nowotwór, powinien wyjaśnić w sposób przejrzysty, dlaczego sugeruje takie właśnie rozpoznanie. Jest to w gruncie rzeczy wyjaśnienie „skojarzone” z zakodowanym wewnątrz systemu pojęciem *nowotwór typu N*. Powinno ono polegać na ukazaniu jakiegoś przekonującego dla człowieka (np. lekarza lub pacjenta) związku między obserwowanymi objawami i sugerowaną diagnozą (w tym: wskazaniu objawów, które są istotne ze względu na diagnozę). Okazuje się, że w przypadku niektórych metod reprezentacji wiedzy (w tym: pojęć) związek taki uzyskuje się w sposób naturalny, w przypadku innych jest to skrajnie trudne. Zagadnienie to omówimy szerzej w rozdziale 5.

4.2. Informatyczne reprezentacje pojęć

Ogólna definicja pojęcia jako *funkcji decyzyjnej* typu $p: X \rightarrow \{0,1\}$ (pojęcia ostre) lub $p: X \rightarrow [0,1]$ (pojęcia rozmyte) dopuszcza wiele możliwych typów reprezentacji wiedzy, które w przypadku konkretnych pojęć pozwalają określić treść formuły p . Warto przypomnieć, że badanie różnych typów reprezentacji stanowi pewne rozszerzenie oryginalnej (czysto ekstensjonalnej) koncepcji Fregego (zob. podrozdział 2.1). Rozszerzenie to nadal jednak pozostaje ujęciem formalnym. Przy takim podejściu rozszerzonym **opisy formalne stosuje się na poziomie reprezentacji pojęciowych (w domyśle: reprezentacji informatycznych)**. Są one konieczne do tego, aby różne czynności dotyczące pojęć – na przykład, rozpoznawanie

⁴⁴ Przez algorytmizację (której warunkiem koniecznym jest formalny zapis pojęć) rozumiemy tutaj opracowanie specjalnych procedur, które pozwalają realizować krok po kroku, w sposób ściśle określony i możliwy do zautomatyzowania, różne procesy operowania pojęciami, w tym nabywanie pojęć, rozpoznawanie obiektów podpadających pod pojęcia i wnioskowanie na podstawie tak a nie inaczej reprezentowanych pojęć. Wybrany aspekt algorytmizacji jest poświęcony podrozdział 4.3.

desygnatów, uczenie się pojęcia czy tworzenie go – można było zalgorytmizować, a następnie zaimplementować w postaci programów komputerowych.

W większości zastosowań informatycznych funkcję p traktuje się jako superpozycję dwóch funkcji c i d , z których pierwsza odpowiada opisowi (klasyfikowanym) obiektów za pomocą ich wyróżnionych cech, druga zaś jest funkcją przyporządkowującą różnym konkretnym opisom decyzje kategoryzacyjne (zero-jedynkowe lub rozmyte). Kwestię tę omówiliśmy bardziej szczegółowo w poprzednim podrozdziale. To właśnie druga z tych funkcji, zakodowana w sposób adekwatny do stosowanej w danym systemie metody reprezentacji wiedzy, jest właściwą reprezentacją pojęcia p .

Zobrazujmy tę charakterystykę ogólną dwoma półformalnymi przykładami konkretnych pojęć: pojęcia *infekcja wirusowa gardła* oraz *niebezpieczne warunki pogodowe w górach*. Pierwsze z nich służy do kategoryzowania stanów chorobowych gardła, które rozróżnia się za pomocą kilku pojęć pokrewnych, typu „infekcja wirusowa”, „infekcja bakteryjna”, „infekcja na podłożu alergicznym”, „lekki stan zapalny” itp. Założmy dalej, że stany podpadające pod te pojęcia opisuje się za pomocą takich przykładowych cech, jak $C_1 = \text{temperatura ciała}$, $C_2 = \text{stopień zaczerwienienia gardła}$, $C_3 = \text{kolor wydzieliny w gardle}$, $C_4 = \text{katar}$ itd... W przypadku ogólnym, abstrahującym od pojęcia i związanych z nim cech, wyróżnia się skończoną liczbę n cech (od C_1 do C_n) oraz skończone zbiory ich wartości v_{ij} (j -ta wartość i -tej cechy). W naszym przykładzie moglibyśmy podać pięć możliwych wartości cechy C_2 (np. *słaby, średni, duży, bardzo duży, ekstremalny*) oraz cztery wartości dla cechy C_3 (np. *biały, żółty, zielony, czerwony*). Opis konkretnego stanu gardła za pomocą określonych w ten sposób wartości, np. (*38, duży, zielony, mały*), odpowiadałby określeniu wartości funkcji c dla konkretnego obiektu x (czyli pewnego stanowi gardła); mówiąc zaś inaczej: konkretnemu opisowi obiektu x w kategoriach cech $C_1, \dots, C_n$. Pojęta ogólnie funkcja c odpowiada oczywiście różnym możliwym opisom tego rodzaju. Funkcja d natomiast odwołuje się do konkretnych wartości cech. Jej formuła określa przyporządkowanie różnym kombinacjom tych wartości konkretnych decyzji kategoryzacyjnych, np. decyzji ostrej typu „Tak, występuje infekcja wirusowa”, albo decyzji nieostrej typu „Infekcja wirusowa występuje ze stopniem pewności 0,7”. Formuła ta jest zakodowana wewnątrz systemu informatycznego, jako pewna reprezentacja pojęciowa, wiążąca w mniej lub bardziej złożony sposób, hipotetyczne wartości cech od C_1 do C_n . Jedną z możliwych reprezentacji są *drzewa decyzyjne* (zob. dalej).

W drugim z zaproponowanych powyżej przykładów, dotyczącym warunków pogodowych, funkcję c moglibyśmy określić za pomocą takich cech jak *intensywność opadów*, *siła wiatru*, *grubość pokrywy śnieżnej* oraz skojarzonych z nimi wartości. Reprezentacją funkcji d mogłaby być natomiast, dla kontrastu z przykładem poprzednim, pewna ustrukturyzowana w drodze uprzedniego treningu sztuczna sieć neuronowa (zob. dalej).

W informatycznej praktyce funkcje d są kodowane za pomocą odpowiednich procedur lub programów komputerowych, których podstawowym celem jest kategoryzacja (w ujęciu praktycznym: rozpoznawanie) obiektów. Programy takie pracują na danych postaci (*cecha*, *wartość*), na przykład (*temperatura ciała*, 39) czy (*stopień zaczerwienienia gardła*, *duży*), które to dane mogą uzyskiwać wprost, za pomocą odpowiednich czujników, lub od człowieka, za pośrednictwem specjalnego interfejsu. Na podstawie konkretnych danych podejmują decyzje o tym, czy dany obiekt należy do zakresu pojęcia, czy nie należy (ewentualnie w jakim stopniu jest to spełnione). Przykładowo: czy lub w jakim stopniu aktualny stan gardła osoby X podpada pod pojęcie *infekcja wirusowa gardła*. Zauważmy wstępnie, że odpowiednie reprezentacje pojęciowe można zakodować w systemie odgórnie (np. zgodnie z wiedzą ekspertów) albo użyć do tego celu pewnego algorytmu ich tworzenia na podstawie dostarczonych przykładów (generować *a posteriori*).

Niezależnie od metody tworzenia, opisywane za pomocą funkcji d **konkretne reprezentacje pojęciowe są zawsze egzemplifikacjami stosowanej w danym systemie metody reprezentacji wiedzy**⁴⁵. W kontekście interesującego nas dalej zagadnienia wyjaśniania metody te warto podzielić na *logicystyczne* i *naturalistyczne*⁴⁶.

Pierwsze z nich, zgodnie z określeniem „logicystyczne”, polegają na zapisie wiedzy za pomocą struktur symboliczno-logicznych, którym w samej logice odpowiadają formuły pewnego rachunku, np. rachunku predykatów pierwszego rzędu. Logicznym „wzorcem” wiedzy pojęciowej jest tutaj pewien *układ implikacji*, w których poprzednikach występują formuły dotyczące wyróżnionych cech obiektów, w następnikach zaś – formuły określające decyzje co do przynależności obiektów do zakresu pojęcia. Forma implikacyjna odpowiada wprost podstawowej funkcji wiedzy pojęciowej, tj. rozpoznawaniu obiektów. Jeśli są spełnione warunki określone w poprzednikach implikacji, obiekt

⁴⁵ Chodzi głównie o systemy sztucznej inteligencji, które są właśnie systemami opartymi na wiedzy (w określony sposób reprezentowanej).

⁴⁶ Podział ten omówiłem szerzej w książce „Umysł a modele maszyn uczących się. Współczesne badania informatyczne w oczach filozofa”, rozróżniając dwa nurty badań nad sztuczną inteligencją, logicystyczny (odwołujący się do pojęć i metod logiki) oraz naturalistyczny (odwołujący się do obserwacji przyrody). Por. (Stacewicz 2010, s. 44-49). Do systemów projektowanych w ramach pierwszego nurtu należą: regułowe systemy eksperckie, systemy do automatyzacji wnioskowań matematycznych (np. dowodzenia twierdzeń) czy systemy przetwarzające język naturalny na podstawie gramatyk typu Chomsky’ego. Do systemów naturalistycznych zaliczyć trzeba sztuczne sieci neuronowe (działające podobnie do biologicznego mózgu) i sztuczne systemy ewoluujące.

W kontekście historycznym warto zauważyć, że omawiany podział ma głębokie korzenie w dwojakim podejściu do modelowania i sztucznej realizacji czynności umysłowych: tym, które wywodzi się z cybernetyki i jest zorientowane naturalistycznie (por. Ashby 1952), i tym, które jest osadzone w logice i logicznych podstawach informatyki (por. (Turing 1936)).

zostaje rozpoznany jako przynależny do pewnej kategorii (jednoznacznie lub w pewnym stopniu). Przykładowo: pojęciu *infekcja wirusowa gardła* mogłaby odpowiadać prosta implikacja typu JEŚLI (*temperatura ciała* > 37) i (*stopień zaczerwienienia gardła* = silny) i (*kolor wydzieliny* = żółty), TO (*infekcja wirusowa* = 1)⁴⁷. Rozpoznanie stanu gardła polegałoby wówczas na prostym sprawdzeniu, czy warunek określony w poprzedniku implikacji (koniunkcja warunków prostszych) jest spełniony⁴⁸.

Wzorzec implikacyjny dopuszcza szereg szczegółowych metod reprezentacji, które rozróżnia się ze względów czysto informatycznych, mając na uwadze pewną całościową formę opisu przetwarzanych danych. Do metod najbardziej typowych należą: *zbiory reguł* w systemach eksperckich (Ignizio 1991), *drzewa decyzyjne* ([Hunt, Marin, Stone 1966]), *tabele decyzyjne* (Pawlak 1991) i *sieci semantyczne* (Flasiński 2011). Metoda ostatnia pozwala osadzić dane pojęcie w całej sieci zależności, która określa relacje między pojęciami, np. relacje nadrzędności i podrzędności (zob. podrozdział 1.2)⁴⁹.

Nie wchodząc w szczegóły wymienionych powyżej metod, można stwierdzić ogólnie, że **ich podstawową zaletą jest możliwość generowania zrozumiałych dla człowieka wyjaśnień, zwana przejrzystością poznawczą w kontekście wyjaśniania**. Zapis wiedzy pojęciowej w postaci struktur sprowadzalnych do (symbolicznych) implikacji lub struktur odwołujących się w sposób istotny do takich implikacji ma dwie zalety. Po pierwsze, jest to możliwość wskazania łańcucha implikacji, których konsekwentne stosowanie w konkretnym przypadku doprowadziło do podjęcia decyzji, że dany obiekt należy do zakresu pojęcia. Po drugie, chodzi o możliwość stwierdzenia, jakie konkretne wartości cech, występujące

⁴⁷ Oczywiście w realnych opisach wskazane pojęcie byłoby reprezentowane przez większą liczbę implikacji, w których poprzednikach mogą występować bardziej skomplikowane formuły niż koniunkcja formuł atomowych.

⁴⁸ Oczywiście w przypadku pojęć i opisów bardziej skomplikowanych, np. za pomocą wielu implikacji, rozpoznanie obiektu (jako desygnatu pojęcia) wymaga przeprowadzenia całego zestawu wnioskowań, a nie sprawdzenia jednego warunku w pojedynczej implikacji.

⁴⁹ W formie przypisu opisujemy dokładniej reprezentację pojęć w postaci drzew decyzyjnych, którą można uznać za najbardziej zaawansowane, ale też często wykorzystywane, rozwinięcie implikacyjnej metody reprezentacji. Pojedyncze drzewo decyzyjne jest grafem skierowanym, który składa się z węzłów połączonych gałęziami. W większości węzłów są sprawdzane pewne warunki, dotyczące pojedynczych cech lub ich grup, np. „jaką wartość ma cecha c_1 ?”. Wychodzące z danego węzła gałęzie reprezentują potencjalne wyniki testów (np. „czy c_1 ma wartość v_{12} ”) i są połączone z węzłami niższego poziomu, składającymi się na kolejny poziom drzewa.

Drzewo rozgałęzia się począwszy od korzenia, w którym sprawdza się pewien warunek początkowy, przez wiele poziomów węzłów testowych, gdzie sprawdza się kolejne warunki, aż do węzłów-liści, które reprezentują decyzje o przynależności bądź braku przynależności danego obiektu do zakresu pojęcia. Dochodzenie do teźe decyzji polega na „przeglądaniu drzewa”, a więc na sprawdzaniu kolejnych warunków skojarzonych z węzłami. Kolejne warunki można traktować jako poprzedniki implikacji, których łańcuch, czyli kolejne gałęzie drzewa, „doprowadza” do finalnej decyzji kategoryzacyjnej (Flasiński 2011).

jawnie w formułach atomowych, były powodem podjęcia takiej decyzji⁵⁰. Zalet tych nie wykazują niektóre przynajmniej reprezentacje naturalistyczne, przede wszystkim zaś koneksyjne.

Reprezentacje logiczystyczne określa się często mianem symbolicznych, jako że **są one sprowadzalne do przejrzystych znaczeniowo formuł logiki formalnej (symbolicznej)**, przede wszystkim implikacji „wiążących” pewne wartości cech. Reprezentacje drugiego rodzaju, czyli naturalistyczne, określa się bardzo często mianem *niesymbolicznych* lub *subsymbolicznych*⁵¹. Ich istota polega na „sięgnięciu” wprost do fizycznych lub biologicznych charakterystyk zjawisk – charakterystyk, które nie zakładają wprost opisu pewnej dziedziny w kategoriach rozróżnialnych cech. Są to zwykle opisy ilościowe (liczbowe), a nie jakościowe. Konkretnie reprezentacje tego typu mają swoje pierwowzory w faktycznych metodach kodowania danych w systemach naturalnych, takich jak mózgi, organizmy żywe lub zbiorowości złożone z takich organizmów. Należą do nich przede wszystkim reprezentacje *koneksyjne* (stosowane w sztucznych sieciach neuronowych), reprezentacje *genetyczne* (wykorzystywane w systemach ewoluujących, opartych np. na algorytmach genetycznych) czy reprezentacje *rojowe* (dostosowane to metod równoległego przetwarzania danych, np. za pomocą algorytmów mrówkowych, zob. (Kari, Rozenberg 2008)). W dalszej analizie skupimy uwagę na metodach pierwszego rodzaju, jako najczęściej bodaj wykorzystywanych w praktyce.

Reprezentacje koneksyjne, czyli metody kodowania pojęć za pomocą *sztucznych sieci neuronowych* (SSN), zawdzięczają swoją nazwę angielskiemu słowu *connections*, które odnosi się do najważniejszego elementu każdej sieci, jakim są połączenia między sztucznymi neuronami. Funkcje konkretnego układu typu SSN zależą, po pierwsze, od **całościowej struktury połączeń** (np. warstwowej), po drugie, **od wag poszczególnych połączeń** (które można interpretować jako ich przepustowości czy siły). Funkcje te są realizowane przez równoległe przetwarzanie sygnału wejściowego sieci, czyli przez równoległą aktywizację neuronów i połączeń między nimi, w wyniku której zostaje wygenerowany sygnał wyjściowy sieci (Tadeusiewicz 1993; Hinton 1989).

Ogólną ideę *reprezentowania pojęć* za pomocą układów typu SSN można przedstawić w trzech punktach: (1) *neurony wejściowe* sieci traktuje się jako elementy rejestrujące wartości pewnych wyróżnionych cech klasyfikowanych

⁵⁰ Choć trzeba przyznać, że w przypadku reprezentacji uwzględniających wiedzę niepewną, opartych np. na logikach wielowartościowych lub rozmytych, zalety te nie są już takie oczywiste. Więcej miejsca poświęcimy temu zagadnieniu w podrozdziale 5.1.

⁵¹ To drugie określenie wskazuje, że reprezentacja naturalistyczna odpowiada czemuś, co tkwi pod poziomem uchwytnych znaczeniowo symboli (jest SUBsymboliczne), na przykład w postaci nieinterpretowalnych wprost parametrów liczbowych. Przedrostek „sub” – w odróżnieniu od „nie” – sugeruje jednak, że istnieje jakiś związek między niższym (subsymbolicznym) poziomem opisu pojęcia i poziomem wyższym (symbolicznym) – związek, który, być może, da się uchwycić algorytmicznie i wykorzystać do generowania zrozumiałych dla człowieka opisów symbolicznych.

obiektów (potencjalnych desygnatów pojęcia), (2) niektóre z *neuronów wyjściowych* wyróżnia się jako elementy generujące sygnały informujące o tym, czy dany obiekt jest desygnatem pojęcia (ewentualnie: w jakim stopniu nim jest), (3) całe *wnętrze sieci*, czyli neurony wewnętrzne i połączenia między nimi, interpretuje się jako system przyporządkowujący wartościom cech obiektów decyzje kategoryzacyjne o przynależności obiektu do zakresu pojęcia. Zgodnie z punktami 1–3 **sieć realizuje konkretną funkcję wielu zmiennych**, $f: C_1 \times C_2 \times \dots \times C_n \rightarrow \{0,1\}$ lub $f: C_1 \times C_2 \times \dots \times C_n \rightarrow [0,1]$, która to funkcja, zgodnie z interpretacją omawianą w rozdziale 2, odpowiada pojęciu (ostremu lub rozmytemu)⁵².

Z powyższego opisu wynika, że właściwą reprezentacją pojęcia w ramach konkretnej sieci, tj. sieci o ustalonych funkcjach neuronów i ustalonej topologii połączeń⁵³, jest **całościowy rozkład wag połączeń międzyneuralnych**. To wagi bowiem, interpretowane jako przepustowości czy siły połączeń, decydują o tym, z jaką siłą sygnały docierają do kolejnych neuronów. Przykładowo: sygnał przechodzący przez dane łącze jest mnożony przez przypisaną doń wagę, co powoduje, że jest on wzmacniany lub osłabiany. Całościowy rozkład wag decyduje o tym, w jaki sposób sygnał wejściowy sieci jest propagowany przez kolejne jej neurony, ostatecznie zaś o tym, z jaką siłą generują swoje sygnały neurony wyjściowe sieci. Matematycznie rzecz biorąc, ów całościowy rozkład jest zbiorem N liczb z pewnego przedziału, gdzie N jest liczbą (zazwyczaj bardzo wielką!) połączeń między neuronami. Mamy zatem do czynienia z **reprezentacją czysto ilościową, niesymboliczną, rozproszoną w ogromnym zbiorze połączeń między węzłami sieci**.

Ze względu na wskazane cechy **reprezentacje konekcyjne nie dostarczają przejrzystych znaczeniowo wyjaśnień, czyli są nieprzejrzyste poznawczo w kontekście wyjaśnień**⁵⁴. Wyjaśnieniem podjętej decyzji kategoryzacyjnej jest zawsze rozproszony układ wag (liczb), które same w sobie, w kontekście decyzji podejmowanych przez sieć, nic nie znaczą. Mówiąc zaś dokładniej, mają „znaczenie” czysto lokalne: decydują o tym, z jaką siłą pewne rozproszone w sieci

⁵² Trzeba zauważyć jednak, że najczęściej pojedyncze sieci neuronowe reprezentują bądź wiele różnych pojęć (wówczas różne neurony wyjściowe lub kombinacje neuronów wyjściowych odpowiadają za rozpoznawanie desygnatów różnych pojęć), bądź realizują bardziej złożone funkcje, które jako jedno z zadań obejmują rozpoznawanie obiektów (nie zawsze interpretowanych jako desygnaty pewnych pojęć).

⁵³ Mówiąc o funkcjach neuronów, mamy na myśli przede wszystkim realizowaną przez nie funkcję wyjścia, czyli funkcję przekształcającą sygnały wejściowe danego neuronu w jego sygnał wyjściowy (dla sieci danego typu jest ona ściśle określona). Mówiąc zaś o topologii połączeń, mamy na myśli całościową strukturę sieci, np. podział sieci na n warstw i istnienie połączeń tylko i wyłącznie między neuronami warstw sąsiadujących (dla sieci konkretnego typu struktura taka jest dokładnie określona). Por. (Tadeusiewicz 1993).

⁵⁴ Nie oznacza to oczywiście, że układy typu SSN są nieprzejrzyste poznawczo w kontekście działania. Ich zasady działania, a także reguły dostosowywania sieci do powierzanych im zadań (uczenia się), są precyzyjnie określone, ale mimo to, nie zapewniają one przejrzystej wiedzy typu „dlaczego” (dlaczego sieć wygenerowała taką a nie inną decyzję),

sygnały dochodzą do poszczególnych neuronów. Całościowa konfiguracja tychże wag również nie niesie ze sobą żadnego bezpośredniego znaczenia. Kształtuje się ona w drodze długotrwałego treningu sieci, który polega na takim dostrajaniu rozproszonych wag międzyneuralnych, aby sieć reagowała poprawnie (np. zgodnie z oczekiwaniami eksperta) na prezentowane jej bodźce, czyli pewne kombinacje wartości wyróżnionych cech. Sama finalna konfiguracja wag nie określa jednak żadnego bezpośrednio zrozumiałego warunku, który wiązałby wspomniane cechy i pozwalał człowiekowi uchwycić istotę reprezentowanego przez sieć pojęcia.

4.3. Od reprezentacji do algorytmów

Specyfika informatycznego podejścia do pojęć polega na tym, że opisane w sposób formalny reprezentacje zbiorów obiektów (reprezentacje pojęciowe) czyni się punktem wyjścia algorytmizacji dotyczących ich procesów. Mówiąc bardzo ogólnie, chodzi o procesy (i problemy zarazem), które wskazaliśmy w podrozdziale 4.1. Są to: *rozpoznawanie* desygnatów pojęcia, *uczenie się* pojęcia na podstawie przykładów, *optymalizacja* istniejącej reprezentacji pojęciowej oraz *tworzenie* pojęcia od podstaw (nazywane też uczeniem się bez nadzoru)⁵⁵. Zagadnieniem odrębnym jest *wyjaśnianie* konkretnych decyzji kategoryzacyjnych, tj. decyzji o przynależności obiektu do zakresu pojęcia. Sposób jego rozwiązania zależy od stopnia przejrzystości poznawczej reprezentacji pojęciowych, której to kwestii poświęcimy rozdział kolejny.

Dla każdego z powyższych problemów **opracowano szereg efektywnych algorytmów, które są dostosowane do przyjętej w danym systemie ogólnej metody reprezentacji wiedzy** (np. regułowej lub konekcyjnej). Przykładowo: w przypadku reprezentacji konekcyjnych, wykorzystywanych wewnątrz sztucznych sieci neuronowych, algorytmizacja uczenia się pojęć polega na precyzyjnym określeniu schematu zmian wag połączeń międzyneuralnych w odpowiedzi na prezentowane sieci przykłady i kontrprzykłady desygnatów. Przy tym dla różnych rodzajów sieci podaje się istotnie różne schematy. Do innych typów reprezentacji, np. zbiorów reguł czy drzew decyzyjnych, schematy konekcyjne nie mają, rzecz jasna, zastosowania, ponieważ konieczne na etapie uczenia się czy tworzenia pojęć zmiany dotyczą zupełnie innych parametrów niż wagi połączeń między sztucznymi neuronami.

Jest to w gruncie rzeczy istotna słabość algorytmicznego podejścia do pojęć: konkretne algorytmy są stosowalne i skuteczne pod warunkiem takiego a nie innego sposobu kodowania pojęć, ten zaś dobiera się do szczególnego narzędzia i szczególnej dziedziny. Przykładowo, regułowy system ekspercki do rozpoznawania chorób krwi wymaga innego typu algorytmów uczenia się i rozpoznawania desygnatów

⁵⁵ Szereg konkretnych algorytmów uczenia się i tworzenia pojęć omówiłem we wcześniejszej swojej monografii, omawiającej m.in. zagadnienie uczenia się pojęć. Zob. (Stacewicz 2010).

pojęć niż sztuczna sieć neuronowa, zaprojektowania np. do klasyfikowania obrazów. Na obecnym etapie badań algorytmizacja ma charakter lokalny i pragmatyczny (ukierunkowany na konkretne zadania), to zaś utrudnia wypracowanie jakiegось jednolitego, uniwersalnego modelu ludzkiej wiedzy pojęciowej. Utrudnia też stworzenie dostatecznie elastycznego systemu informatycznego, który korzystałby z pojęć bardzo różnorodnych, przynależnych do różnych dziedzin.

Wybór konkretnej reprezentacji wiedzy powoduje wpływa nadto na łatwość, a niekiedy nawet na możliwość, algorytmicznego rozwiązania wskazanych powyżej typowych problemów. Przykładowo, reprezentacje logicystyczne nadają się znakomicie do algorytmizacji problemu wyjaśniania (dlaczego została podjęta taka a taka decyzja kategoryzacyjna), ale nie są już tak skuteczne w przypadku uczenia się pojęć (zwłaszcza wtedy, gdy uczenie się ma zachodzić bez nadzoru). Reprezentacje naturalistyczne z kolei, w tym koneksyjne, prowadzą do efektywnych algorytmów uczenia się i tworzenia pojęć, nie zapewniają jednak, ze względu na swoją nieprzejrzyistość poznawczą, równie skutecznej strategii wyjaśniania.

4.4. Podsumowanie

1. Informatyczne badania nad pojęciami sytuują się głównie w obszarze **sztucznej inteligencji**, czyli w tym dziale informatyki, który ma na celu algorytmizację i automatyzację ludzkich czynności poznawczych (w tym: czynności posługiwania się pojęciami).

2. W informatyce stawia się pięć ogólnych zagadnień dotyczących pojęć rozumianych jako funkcje decyzyjne: a) zagadnienie **reprezentowania** pojęć (w ramach ogólniejszych metod reprezentacji wiedzy), b) zagadnienie **rozpoznawania** desygnatów pojęć; c) zagadnienie **uczenia się** pojęć na podstawie przykładowych opisów desygnatów; d) zagadnienie **tworzenia** pojęć (bez przykładowych opisów desygnatów), e) zagadnienie **wyjaśniania** decyzji kategoryzacyjnych.

3. Informatyczne reprezentacje pojęciowe można podzielić na **logicystyczne** (symboliczne i wywodzące się z logiki) oraz **naturalistyczne** (subsymboliczne i wywodzące się z obserwacji natury). Typowe przykłady pierwszych to zbiory reguł, drzewa decyzyjne i tabele decyzyjne; typowe przykłady drugich to struktury koneksyjne, czyli rozkłady wag połączeń w sztucznych sieciach neuronowych.

4. Do konkretnych reprezentacji pojęciowych dobiera się **algorytmy** pozwalające tworzyć i wykorzystywać pojęcia, a także wyjaśniać powody podejmowanych decyzji kategoryzacyjnych.

Rozdział 5

Zagadnienie przejrzystości poznawczej reprezentacji pojęciowych

W tym rozdziale będziemy nadal nawiązywać do informatyki, wskazując na pewien problem, który wiąże się z próbami obliczeniowego modelowania i/lub sztucznego realizowania pojęć. Jest to problem przejrzystości poznawczej reprezentacji pojęciowych, zwany obrazowo problemem *czarnej skrzynki*. Ma on wymiar znacznie szerszy niż poznawcza dostępność samej wiedzy pojęciowej, na co zwrócimy uwagę w podrozdziale 5.1. W odniesieniu do pojęć natomiast pojawia się on na dwóch poziomach. Po pierwsze, na poziomie samej reprezentacji pojęciowej, a więc – ujmując rzecz matematycznie – formuły funkcyjnej, która odpowiada pojęciu. Po drugie zaś, dotyczy on struktury wyjaśnień uzasadniających konkretne decyzje kategoryzacyjne, podejmowane przez ludzi lub systemy informatyczne.

5.1. Metafora i problem czarnej skrzynki

Porównywanie systemów operujących pojęciami (w tym: rozpoznających obiekty) do czarnych skrzynek wpisuje się dobrze w czysto ekstensjonalną strategię opisu pojęć jako funkcji typu $f: U \rightarrow \{0,1\}$ lub $f: U \rightarrow [0,1]$ o nieokreślonych formułach funkcji f ⁵⁶. W ramach takiej strategii analizuje się ekstensje pojęć i prowadzące do nich przyporządkowania, a nie reprezentowane w konkretny sposób intensje, czyli znaczenia. Reprezentacje pozostają „ukryte”, a to sprawia, że układ operujący pojęciami redukuje się do „czarnej skrzynki”, która przekształca dane wejściowe w decyzje kategoryzacyjne. Podejście tego rodzaju motywuje się najczęściej potrzebą analiz czysto matematycznych (abstrakcyjnych), niekiedy jednak **niemożnością**

⁵⁶ Przypomnijmy, że jest to zgodne zarówno z oryginalnym podejściem Fregego (zob. podrozdział 2.1), jak i ogólnym ujęciem informatycznym (zob. rozdział 4).

wiarygodnego opisu wnętrza pojęciowej czarnej skrzynki⁵⁷. W dalszej części podrozdziału omówimy szersze tło tej drugiej ewentalności.

Koncepcję *czarnej skrzynki* wprowadzili do nauki cybernetycy i psychologowie behawioryści, zajmujący się odpowiednio: ogólną teorią sterowania i komunikacji w układach przetwarzających informacje oraz teorią zachowania zwierząt i ludzi. Ci drudzy podnosili szczególnie mocno problem zasadności pewnych tradycyjnych metod badania umysłu, takich jak introspekcja i wieńczący ją werbalny opis postrzeganych świadomie stanów umysłowych (Skinner 1938). Metody te uznawali za mało wiarygodne, ponieważ ich wyniki są zawsze obciążone czynnikami subiektywnymi, a przez to niemierzalnymi, takimi jak nastawienia, oczekiwania czy chwilowe nastroje (a mówiąc ogólniej: cały historyczny bagaż świadomych i nieświadomych przeżyć konkretnego człowieka). Metafora czarnej skrzynki miała zredukować umysł do tego, co dobrze opisywalne, mierzalne i zewnętrzne, a więc do układu relacji między dochodzącymi do organizmu bodźcami i obserwowalnymi reakcjami. Przykładowo: zamiast posługiwać się nieprecyzyjnym i subiektywnym opisem przeżycia typu „*X* odczuwa silny ból” behawiorysta użyłby sformułowań zdających sprawę z wyników konkretnych eksperymentów – sformułowań typu „Przy takim a takim natężeniu bodźca *B*, zarejestrowano taki a taki zestaw mierzalnych reakcji organizmu *X*-a”⁵⁸. Intencja behawioralnej redukcji była proekspłanacyjna: wyjaśnienia w kategoriach postrzeganych introspekcyjnie stanów uznawano za słabsze (mniej naukowe) niż wyjaśnienia w kategoriach korelacji między układami mierzalnych bodźców i reakcji. Wyróżniając określone typy bodźców i reakcji, a także postulując pewien matematyczny opis relacji między nimi, np. układ implikacji, behawioryści przekształcali metaforę w model (nie wychodząc jednak poza postulat redukcji opisu do tego, co fizyczne i mierzalne). Cybernetyka zaś stosowała tego rodzaju charakterystyki i modele na poziomie bardziej ogólnym: nie tylko do opisu ludzi czy zwierząt, lecz do opisu wszelkich systemów przetwarzających informacje i wchodzących w interakcję ze swoim środowiskiem (w tym: sztucznych; por. (Ashby 1961)). Warto podkreślić, że przy obydwu podejściach, behawioralnym i cybernetycznym, mianem czarnej skrzynki określano obiekt modelowany (np. umysł lub organizm), a nie sam model. Preferowana przez cybernetyków metoda czarnej skrzynki polegała na do-

⁵⁷ Sformułowanie „czarna skrzynka” ma charakter metaforyczny. Chodzi tu o metaforę zaglądania do wnętrza pewnego artefaktu (najczęściej: modelu lub pewnego wstępnego konstruktu pojęciowego, będącego punktem wyjścia modelu), które to wnętrze pozostaje ukryte lub trudno dostępne, czyli poznawczo nieprzejrzyste.

⁵⁸ Filozoficznym odpowiednikiem podejścia behawioralnego w psychologii był program behawioryzmu logicznego w filozofii umysłu. Program ten postulował opis zjawisk umysłowych bez używania jakichkolwiek predykatów „mentalnych”, odnoszących się do wewnętrznych (mentalnych czy świadomych) stanów podmiotu. Predykaty te traktowano co najwyżej jako pewne określenia skrótowe, które należy rozwinać, a tym samym precyzyjnie określić, za pomocą zdań o znaczeniu czysto fizycznym, dotyczących zatem obserwowalnych i mierzalnych stanów, zdarzeń czy procesów (por. (Ryle 1970)).

chodzeniu do modelu przez rejestrowanie i analizowanie zewnętrznych zachowań obiektu modelowanego⁵⁹.

Wraz z rozwojem technik algorytmicznych – przynależnych już do informatyki, a nie cybernetyki – eksplanacyjny potencjał pojęcia i metody czarnej skrzynki zaczął się wyczerpywać. Wyjaśnienia w postaci „zewnętrznych” korelacji między wielkościami rejestrowanymi na wyjściu i wejściu przestały wystarczać. Metody algorytmiczne, a zwłaszcza symboliczne, znane z badań nad sztuczną inteligencją, **zapewniały cenny poznawczo wgląd w możliwe związki między wejściem a wyjściem**. Przykładowo: jeśli były to metody oparte na logice, pozwalały ujawnić przejrzyste znaczeniowo łańcuch wnioskowań prowadzących od wejściowych przesłanek do wyniku. Mówiąc zaś ogólniej, sam algorytm generowania wyników, przez swą intersubiektywnie dostępną strukturę, pozwalał określić (a w pewnych przypadkach: zrozumieć) pewne istotne elementy relacji wejście-wyjście.

W takiej sytuacji określenie „czarna skrzynka” stało się synonimem układu, który generuje poprawne (lub: akceptowalne) wyniki, ma też pewną intersubiektywnie dostępną „zawartość”, nie zapewnia jednak (i nie przedstawia) zrozumiałych dla człowieka wyjaśnień uzyskiwanych wyników. Analizując taki układ, możemy uzyskać konkretną wiedzę „jak” (np. jak dochodzi do podjęcia decyzji czy rozwiązania problemu), nie możemy jednak zdobyć odpowiednio czytelnej wiedzy „dlaczego”. W przypadku układów tego rodzaju określenie „czarna skrzynka” dotyczy zatem struktury wyjaśnień, a nie struktury relacji wiążącej wejście z wyjściem. Choć ta druga jest znana, to sama w sobie nie prowadzi do uchwytynego dla człowieka (również konstruktora czy programisty) układu wyjaśnień⁶⁰.

Dobrym przykładem czarnych skrzynek w sensie algorytmicznym i eksplanacyjnym zarazem (nie zaś cybernetycznym czy behawioralnym) są *sztuczne sieci neuronowe* (zob. podrozdział 4.2). Algorytmy kontrolujące ich sposób działania, w tym uczenia się, są precyzyjnie określone, a co za tym idzie

⁵⁹ Rozumowanie w kategoriach czarnej skrzynki znajdujemy w idei słynnego testu Turinga, który dotyczy stwierdzania hipotetycznej inteligencji maszyn na podstawie oceny werbalnych zachowań tychże (Turing 1950). Zgodnie z tą ideą maszynę należy uznać za inteligentną wówczas, gdy generowane przez nią odpowiedzi na pytania są w dostatecznie dużym stopniu nieodróżnialne od wypowiedzi ludzi. O behawioralnym charakterze testu decyduje fakt, że porównując maszynę z człowiekiem abstrahuje się z jednej strony, od jakichkolwiek jego przeżyć czy stanów świadomych, z drugiej strony zaś, od znajomości jakichkolwiek stanów wewnętrznych maszyny. Ocenia się tylko i wyłącznie zewnętrzne wyniki tych stanów, czyli obserwowalne zachowania językowe.

⁶⁰ Warto dopowiedzieć, że kwestia wyjaśnialności i zrozumiałości generowanych przez system wyjaśnień jest względna. Znaczy to, że zależnie od oczekiwań danej grupy osób (tzw. interesariuszy) wymagane są inne typy wyjaśnień. Przykładowo: programiści są zainteresowani czytelnością układu instrukcji programu i znajomością ogólnego algorytmu, który program realizuje (zakłada się przy tym, że znają oni określony język programowania), użytkownicy natomiast oczekują wyjaśnień w kategoriach zależności między obiektami z dziedziny, której dotyczy problem rozwiązywany przez dany system (np. figurami geometrycznymi, jeśli program rozwiązuje problem geometryczny). Por. (Zednik 2019).

całkowicie przejrzyste co do swojej struktury (przynajmniej dla osób zajmujących się profesjonalnie takimi układami). Ponadto, na każdym etapie działania sieci istnieje możliwość dokładnego odczytu konkretnych wartości wag międzyneuronalnych i sygnałów wyjściowych poszczególnych neuronów. Mimo to, wiedza płynąca ze znajomości wspomnianych algorytmów i odczytów nie przekłada się wprost na wysokopoziomowe uzasadnienia, które wyjaśniałyby, w języku stosowanym do opisu powierzonego sieci zadania, dlaczego w konkretnym przypadku sieć dochodzi do takiego a nie innego wyniku. Przykładowo: pewna sieć wspomagająca lekarzy może rozpoznawać prawidłowo choroby, czyli generować prawidłowe diagnozy na podstawie przedstawianych jej objawów. Mimo to, ani znajomość jej struktury wewnętrznej (rozkładu i wartości wag), ani znajomość reguł przetwarzania sygnałów przez pojedyncze neurony, ani znajomość algorytmu zmiany wag podczas wcześniejszego treningu sieci, nie pozwalają sformułować wprost, bez użycia specjalnych algorytmów „interpretujących”, uzasadnienia diagnozy w języku zrozumiałym dla lekarza czy pacjenta (języku objawów, jednostek chorobowych, związków przyczynowo-skutkowych między objawami itp.).

Mając na uwadze powyższy przykład, możemy objaśnić w sposób ogólny sens przesunięcia znaczeniowego, jakie dokonało się w sposobie rozumienia terminu „czarna skrzynka”. W ujęciu tradycyjnym, cybernetyczno-behawioralnym, za czarne skrzynki uznawano układy o nieprzejrzystej poznawczo strukturalnie wewnętrznej. Wskutek tejże nieprzejrzystości układy takie opisywano tylko i wyłącznie w kategoriach korelacji między danymi wejściowymi (bodźcami) i wyjściowymi (reakcjami) – uznając takie opisy za jedyne wartościowe naukowo wyjaśnienia tego, jak dany układ działa. W ujęciu nowszym, zgodnym ze współczesnymi dokonaniem informatyki, **znaczeniowy punkt ciężkości określenia „czarna skrzynka” przesunął się w kierunku siły i struktury wyjaśnień**. Chodzi jednak o wyjaśnienia, które dotyczą nie sposobu działania systemu, lecz generowanych przezeń wyników (por. Creel 2020). Rozpiszmy tę obserwację dokładniej.

O systemach informatycznych, czyli działających na podstawie pewnych algorytmów, nie można twierdzić zasadnie, że przypominają czarne skrzynki w sensie cybernetyczno-behawioralnym. Z zasady bowiem znamy strukturę wchodzącego w grę algorytmu i odpowiadającego mu programu (nawet jeśli ma ona ogromną złożoność), możemy także manipulować wewnętrznymi parametrami algorytmu i obserwować adekwatne do tych manipulacji zmiany w zachowaniu sterowanego algorytmicznie systemu⁶¹. Pozostaje to prawdą również w odniesieniu do takich

⁶¹ Z ogólniejszego punktu widzenia zaletą metody algorytmicznej – w odróżnieniu od metod angażujących niewerbalizowaną do końca intuicję – jest właśnie intersubiektywna dostępność i kontrolowalność schematu, według którego postępujemy. Te cechy powodują, że metodę algorytmiczną można łatwo zautomatyzować, np. za pomocą komputerów cyfrowych (Stacewicz 2016).

systemów, których oprogramowanie powstaje lub podlega zmianom w drodze uczenia się, a więc dzięki zastosowaniu pewnego programu wyższego poziomu odpowiedzialnego za zmianę parametrów programu właściwego. W przypadku systemów uczących się również znamy właściwości obydwu algorytmów/programów, możemy dokonywać zmian ich wewnętrznych parametrów oraz obserwować efekty tych zmian. Ze strukturalnego i operacyjnego punktu widzenia nie są to zatem żadne nieprzejrzyste dla programisty (czy nawet: użytkownika) czarne skrzynki. Mimo to, z perspektywy oczekiwanych przez użytkownika wyjaśnień podejmowanych decyzji, systemy takie okazują się nieskuteczne, to znaczy ich sposób działania (a niekiedy też: uczenia się) nie zapewnia przejrzystej struktury wyjaśnień. I w tym właśnie sensie, sensie eksplanacyjnym, przyrównanie takich układów do czarnych skrzynek wydaje się zasadne. Zob. też (Skowron 2021, s. 118).

Nieprzejrzystość poznawczą w sensie eksplanacyjnym przypisuje się najczęściej informatycznym *systemom uczącym się*, w tym sztucznym sieciom neuronowym (Zednik 2019). Oprogramowanie do automatycznego uczenia się nie jest jednak jedynym czynnikiem, który może prowadzić do efektu nieprzejrzystości. Co więcej, niektóre metody uczenia się, o ile są dostosowane do symbolicznych metod reprezentacji wiedzy, efektu tego nie powodują. Zestawmy w sposób niewyczerpujący różne możliwe powody tej niepożądaną w wielu sytuacjach właściwości.

1. *Złożoność strukturalna*. Zbyt duża złożoność układu przetwarzającego dane (np. ogromna liczba reguł w systemie eksperckim lub połączeń między neuronami w sieci neuropodobnej) może powodować trudności z wyodrębnieniem tych elementów, które wpłynęły istotnie na wygenerowanie konkretnego wyniku. Ponadto, jeśli układ wchodzi w dynamiczne interakcje ze swoim środowiskiem, w wyniku czego jego struktura wewnętrzna zmienia się, to identyfikacja elementów istotnych, dokonywana *post factum*, może być w ogóle nieosiągalna (Creel 2020).

2. *Uczenie się*. Jeśli system informatyczny doskonali swoje działanie (a niekiedy: jest tworzony od podstaw) na podstawie pewnego algorytmu uczenia się, to algorytm ten może powodować dodawanie do programu sterującego systemem pewnych technicznych, trudnointerpretowalnych parametrów, które mają na celu tylko i wyłącznie efektywne działanie systemu (np. dopasowanie jego sposobu działania do pewnych przykładowych par [dane, oczekiwany wynik]).

Efekt nieprzejrzystości jest tym większy, im szerzej na etapie uczenia się są wykorzystywane wybory losowe. Ich losowy charakter sprawia, że finalna struktura systemu kształtuje się w sposób nieprzewidywalny, a to powoduje, że nawet twórca systemu może nie rozumieć wpływu poszczególnych elementów na generowane przez system wyniki (Zednik 2019).

3. *Naśladowanie natury.* Za nieprzejrzystość systemu może odpowiadać fakt, że tworzy się go na wzór pewnych nie dość dobrze poznanych układów naturalnych, jak komórki biologiczne, mózgi, ewoluujące gatunki itp. – przyjmując po prostu, że układy takie działają w przyrodzie wystarczająco efektywnie. Ponieważ działanie tego typu układów (w tym: ludzkiego mózgu), a także wpływ ich wewnętrznych zmian na obserwowane zachowania, wciąż jest przedmiotem badań naukowych i daleko nam do pełnego zrozumienia wielu zależności, to systemy konstruowane na podobieństwo układów naturalnych dziedziczą niejako pierwotną nieprzejrzystość tych ostatnich. Obserwacja ta dotyczy, na przykład, pewnych typów sztucznych sieci neuronowych (Hinton 1989).

4. *Operowanie na danych niepewnych.* Jeśli system operuje na danych, które są obciążone pewnymi stopniami niepewności, to procedura dochodzenia do finalnej decyzji, nawet w ramach symbolicznych metod reprezentacji wiedzy, jest bardzo nieintuicyjna. Stopnie (nie)pewności wpływają na decyzje cząstkowe w sposób nieoczywisty – np. zgodnie z arbitralnymi regułami logiki rozmytej (Klir, Yuan 1995). Wyjaśnienie ostatecznej decyzji, wybieranej spośród wielu hipotetycznych rozwiązań o różnych stopniach pewności, zależy od tego, jaki mechanizm propagacji niepewności (statystyczny, oparty na logice wielowartościowej, oparty na logice rozmytej...) zastosowano. Wobec możliwości wyboru wielu różnych mechanizmów wyjaśnienie ostatecznej decyzji nie jest jednoznaczne.

5. *Brak procedur translacji.* W przypadku wielu systemów (w tym: sztucznych sieci neuronowych) nie mamy do dyspozycji wiarygodnych metod translacji reguł niskiego poziomu (tworzonych często w drodze uczenia się) na przejrzyste znaczeniowo reguły symboliczne. Sytuacja taka występuje wówczas, gdy uczenie się przebiega tylko i wyłącznie na poziomie operacyjnym, nie angażuje żadnych struktur symbolicznych wyższego poziomu, a polega na „dopasowywaniu” struktury operacyjnej (np. wag połączeń międzyneuralnych wewnątrz sieci neuronowej) do decyzji wymuszanych przez nauczyciela lub środowisko⁶².

Spośród wskazanych powyżej czynników powodujących nieprzejrzystość poznawczą systemów operujących pojęciami za najważniejsze uznajemy **złożoność strukturalną oraz brak efektywnych reguł przejścia z poziomu reguł**

⁶² Przykładowo: system uczy się w taki sposób, aby odpowiadać określonymi sygnałami na pewne przykładowe wzorce (np. obrazy czy dźwięki) oraz podobnymi sygnałami na podobne wzorce; po zakończeniu treningu system reaguje prawidłowo, nie dostarcza jednak żadnych wyjaśnień symbolicznych.

operacyjnych (gwarantujących sprawne działanie systemu) **na poziom dostępnych dla użytkownika wyjaśnień**. Czynniki te są ze sobą powiązane, ponieważ odpowiednio złożone systemy informatyczne są zorganizowane, a także projektowane „warstwowo”, począwszy od warstw kodu maszynowego (reguł bardzo niskiego poziomu) aż do warstwy dostępnego dla człowieka interfejsu (komunikacyjnego lub programistycznego). Złożoność systemu jest, z jednej strony, pochodną interakcji między kolejnymi warstwami, z drugiej strony zaś, może zostać zniwelowana przez odpowiednio czytelne struktury i wyjaśnienia na warstwach wysokiego poziomu. Problem w tym, że w przypadku niektórych systemów, zwłaszcza uczących się, niektóre parametry odpowiedzialne za skuteczność systemu są generowane bezpośrednio na warstwach najniższych (bo np. tych warstw dotyczy uczenie się) i trudno je zinterpretować na warstwach wyższych.

5.2. Przejrzystość poznawcza reprezentacji pojęciowych

W rozdziale 1, omawiając zagadnienie wiedzy i konstytuujących ją pojęć, posługiwaliśmy się rozróżnieniem pojęć w sensie *logicznym* i pojęć w sensie *psychologicznym*. Te pierwsze – rozumiane jako znaczenia nazw generalnych, wyrażalne w pewnym intersubiektywnie dostępnym języku i precyzowane za pomocą ścisłych definicji – są w sposób oczywisty przejrzyste poznawczo. Ich przejrzystość, a mówiąc dokładniej przejrzystość odpowiednich reprezentacji, polega na określeniu ścisłych warunków, które muszą spełniać desygnaty danej nazwy. W przypadku zaawansowanych pojęć naukowych warunki te mogą być trudno zrozumiałe dla niespecjalistów; jednak, przy założeniu znajomości odpowiedniej dziedziny, są one jasno interpretowalne.

Jeśli chodzi o pojęcia w sensie *psychologicznym* – rozumiane jako pewne nie-naoczne przedstawienia lub uogólnione schematy pamięciowe zbiorów obiektów – to w wielu sytuacjach ich przejrzystość poznawcza stoi pod dużym znakiem zapytania. Pozostaje to prawdą, nawet jeśli osobie operującej danym pojęciem wydaje się, że ma absolutny wgląd w znaczenia używanych słów czy uzasadnienia swoich decyzji kategoryzacyjnych⁶³. Nieprzejrzystość poznawcza pojawia się zazwyczaj wtedy, kiedy pewna osoba przyswoiła sobie pojęcie w sposób praktyczny, na przykład na podstawie przedstawianych jej przykładów czy na podstawie samodzielnych prób kategoryzacji. Osoba ta może dysponować pewną wiedzą

⁶³ Pewna forma nieprzejrzystości poznawczej, wyrażająca się w braku pewności podmiotu poznawczego co do przynależności obiektów do zakresu danej nazwy, jest przyczyną tego, że psychologowie wyróżniają (obok jednoznacznie interpretowalnych reprezentacji klasycznych) reprezentacje prototypowe i egzemplarzowe (zob. rozdz. 1.4), zaś informatycy formalizują niektóre pojęcia (pojęcia nieostre) za pomocą teorii zbiorów rozmytych (zob. podrozdział 3.2).

praktyczną, potrafi zatem odróżniać trafnie desygnaty pojęcia od niedesygnatów, ale nie zna dokładnie reguły (warunku), która kieruje jej decyzjami kategoryzacyjnymi.

Omawiane w podrozdziale 4.2 informatyczne reprezentacje pojęciowe pozwalają uzyskać pewien wgląd w zagadnienie nieprzejrzystości pojęć rozumianych psychologicznie. Jest to wgląd za pośrednictwem modelu. Narzędziem modelowania jest pewien system, np. system sztucznej inteligencji, który uznaje się wstępnie za teoretyczny odpowiednik pewnego zespołu funkcji poznawczych, np. funkcji percepcyjnych czy powiązanych z wnioskowaniem. Jak wyjaśnialiśmy w podrozdziale 5.1, system taki może być, mimo intersubiektywnie dostępnej struktury wewnętrznej, *poznawczo nieprzejrzysty* (w sensie eksplanacyjnym). Nie znaczy to oczywiście, że systemu takiego nie można traktować jako wartościowego modelu. O ile dostarcza on dobrych przewidywań na poziomie zachowania (układu modelowanego), to sama intersubiektywna dostępność i kontrolowalność sterującego nim programu, czyni zeń dobry model.

System taki, rozumiany właśnie jako dobry model, może dostarczyć wyjaśnienia nieprzejrzystości poznawczej w układzie modelowanym, również w odniesieniu do pojęć rozumianych psychologicznie. Możliwe do zidentyfikowania powody nieprzejrzystości na poziomie modelu można potraktować jako wyjaśnienie nieprzejrzystości w układzie modelowanym. Dotyczy to także reprezentacji pojęciowych, które muszą przecież wchodzić w skład systemów modelujących wszelkie złożone czynności poznawcze. Jeśli ogólny model jest adekwatny, to powody nieprzejrzystości użytych w nim reprezentacji (informatycznych) można uznać za wyjaśnienie nieprzejrzystości reprezentacji realnych, np. psychologicznych. Przypomnijmy, że niektóre przynajmniej z tych powodów wskazaliśmy w podrozdziale 5.1.

Gdy reprezentacje informatyczne są wykorzystywane poza kontekstem modelowania, w ramach zadań czysto praktycznych (np. inżynierskich, jak sterowanie), to problem ich nieprzejrzystości nie dotyczy w ogóle umysłu i sposobów kodowania wiedzy w umyśle. Dotyczy on **zdolności danego systemu do wyjaśniania generowanych wyników, w tym podejmowanych decyzji**. Choć wyjaśnienia te powinny być dostępne i przekonujące dla ludzkiego umysłu, to w żaden sposób nie odnoszą się do struktur kognitywnych w ludzkim umyśle. Na pewno zaś nie służą one wyjaśnianiu, dlaczego niektóre z tych struktur są poznawczo nieprzejrzyste.

Ogólne powody, dla których pewne informatyczne reprezentacje pojęciowe – rozważane zarówno w kontekście modelowania, jak i poza nim – cechuje nieprzejrzystość, zostały wskazane (niewyczerpująco) w podrozdziale poprzednim. Do pięciu omówionych tam kwestii, tj. złożoności strukturalnej, uczenia się, naśladowania natury, operowania na danych niepewnych i braku procedur translacji, warto dodać szóstą.

Mówiąc ogólnie, **polega ona na „uwikłaniu” danej reprezentacji pojęciowej w całościową reprezentację wiedzy, na podstawie której działa system.**

System taki podejmuje trafne decyzje kategoryzacyjne, decyzje te dotyczą obiektów bardzo zróżnicowanych, ale analiza procesu decyzyjnego nie pozwala w sposób czytelny wyodrębnić osobnych reprezentacji, które „są uaktywniane” w przypadku decyzji konkretnego typu. W dochodzeniu do nich uczestniczy bowiem bardzo wiele elementów systemu. Ponownie, idealnym przykładem są tu sztuczne sieci neuronowe, w których wiedza jest rozproszona między ogromną liczbę wag połączeń, a finalne decyzje zależą od całościowego układu tych wag, nie zaś od ich konkretnych podzbiorów. Mówiąc obrazowo, w przypadku wielu sieci reprezentacja danego pojęcia jest rozproszona i „przemieszana” z reprezentacjami innych pojęć.

W podrozdziale 4.2, dotyczącym metod informatycznych, rozróżniliśmy *reprezentacje symboliczne*, takie jak drzewa decyzyjne i sieci semantyczne, oraz *reprezentacje niesymboliczne*, takie jak sztuczne sieci neuronowe czy zapisy genetyczne. Te pierwsze, cechuje **znaczny stopień przejrzystości**⁶⁴, te drugie, ze względu na niskopoziomowy charakter zapisu, **są poznawczo nieprzejrzyste**. Wracając do pozainformatycznego, a dotyczącego ludzkiej wiedzy, rozróżnienia między pojęciami w sensie logicznym i w sensie psychologicznym, warto podkreślić, że reprezentacje informatyczne obydwu typów (i symboliczne, i niesymboliczne) mogą być postrzegane jako ogniwo pośrednie między psychologicznym i logicznym sposobem rozumienia pojęć. Z jednej strony, **są określone formalnie, w sposób pozwalający osadzić je w różnego rodzaju algorytmach** – i to zbliża je do pojęć w sensie logicznym. Z drugiej strony, **są mocno zindywidualizowane, czemu sprzyja fakt, że mogą powstawać w wyniku (automatycznego) uczenia się**. To zaś czyni je czymś pokrewnym pojęciom w sensie psychologicznym – w razie potrzeby zaś, pozwala wykorzystywać je w *stricte* psychologicznych modelach funkcji poznawczych, np. modelach pamięci czy rozumowania⁶⁵.

O tym, czy informatyczne reprezentacje pojęciowe pewnego typu są przydatne bardziej do zapisu pojęć w sensie logicznym, czy w sensie psychologicznym decyduje stopień ich przejrzystości poznawczej (na poziomie wyjaśnień). Im jest on większy, tym bardziej dany typ reprezentacji nadaje się do formalnego opisu pojęcia w sensie logicznym. Przykładowo, implikacje zawarte w klasycznym drzewie decyzyjnym są tak precyzyjne i czytelne, że można je potraktować zbiorczo jako definicję konstytuującą takie a takie pojęcie w sensie logicznym.

Co ważne i ciekawe, informatyka dostarcza pewnych **narzędzi zwiększania przejrzystości, a więc wydobywania istotnej i czytelnej wiedzy z zapisów niesymbolicznych lub zapisów symbolicznych o dużym stopniu złożoności**.

⁶⁴ Choć należy zauważyć, że w przypadku dużej złożoności odpowiedniej reprezentacji (np. dużego rozmiaru drzewa decyzyjnego) przejrzystość może być ograniczona; jej uzyskanie może też wymagać użycia specjalnych procedur kompresji wiedzy.

⁶⁵ Przy czym modele te mogą dotyczyć różnych poziomów systemu poznawczego, nie tylko poziomu *stricte* psychologicznego. Na przykład, sztuczne sieci neuronowe, o ile używa się ich w kontekście modelowania czynności umysłowych, dotyczą neurobiologicznego poziomu reprezentacji.

Metody takie istnieją, choć w obecnej pracy ich nie opisywaliśmy; w tej sprawie zob. np. (Zednik 2019). Przynajmniej w niektórych sytuacjach, gdy obiekty informatyczne pełnią funkcję modeli struktur wewnątrzmysłowych, metody te można potraktować jako efektywne narzędzie dochodzenia do pojęć w sensie logicznym na podstawie analizy pojęć w sensie psychologicznym. Ze względu na „punkt dojścia” analiza ta musi brać pod uwagę nie tylko to, jak faktycznie ludzie posługują się pojęciami (w tym: jak reprezentują zbiory obiektów w swojej pamięci), ale również to, na ile ich czynności spełniają pewne normy. Przykładowo: próbując wyabstrahować z reprezentacji konekcyjnej jednoznaczny, definicyjny, opis pewnej kategorii obiektów, należy nie tylko „wydobyć z sieci” uogólniony opis tej kategorii, ale nadto dostosować ten opis do znanych z logiki reguł poprawnego definiowania.

Przechodząc na koniec do praktyki użytkowania sztucznych systemów opartych na wiedzy, warto postawić pytanie, na ile przejrzystość poznawcza kodowanych wewnątrz nich reprezentacji pojęciowych jest pożądana z punktu widzenia użytkownika lub konstruktora systemu. Twierdzimy, że **przejrzystość odgrywa tym większą rolę, im bardziej złożone jest pojęcie**. Proste decyzje kategoryzacyjne (np. „*ten przedmiot jest czerwony*”) nie wymagają specjalnego wglądu w strukturę reprezentacji: decyzja może tu zapadać „bezrefleksyjnie”, w wyniku reakcji, która w razie konieczności jest utrwalana podczas prostego procesu uczenia się (polegającego np. na odróżnianiu obiektów czerwonych od nieczerwonych). Pojęcia złożone natomiast (np. *infekcja wirusowa gardła*) wymagają reprezentacji na tyle przejrzystej, aby odwołując się do niej, można było uzyskać wyjaśnienia konkretnych decyzji (np. dlaczego ten oto konkretny stan gardła podpada pod pojęcie takiej a takiej infekcji). Wyjaśnień oczekują zarówno użytkownicy, nierozumiejący pojęcia złożonego wprost i bezpośrednio, jak i programiści, odpowiedzialni za doskonalenie systemu (m.in. na podstawie generowanych przezeń uzasadnień). Złożoność pojęcia jest tutaj zarówno powodem dążenia do maksymalnej przejrzystości, jak i pewną przeszkodą. Przeszkoda polega na tym, że układy o dużej złożoności strukturalnej (a tej sprzyja złożoność kodowanych wewnątrz nich reprezentacji pojęciowych) przypominają czarne skrzynki, które trudno jest uczynić przejrzystymi. Kwestię tę omawialiśmy w podrozdziale poprzednim.

5.3. Różne typy i poziomy wyjaśnień

Rozpatrując zagadnienie wyjaśnialności decyzji podejmowanych przez systemy operujące pojęciami, należy mieć na uwadze, że systemy te – zarówno naturalne, jak i sztuczne – mają *strukturę wielopoziomową*. W przypadku człowieka, na przykład, mamy do czynienia z dość dobrze opisaną przez nauki szczegółowe

sekwencją coraz bardziej złożonych, wewnątrz- i zewnątrzmysłowych struktur. Są to między innymi: struktury neurobiologiczne (np. neurony, połączenia między nimi, grupy neuronów), poznawcze (np. pamięć, uwaga i wyobrażenia) i społeczne (np. grupa etniczna czy zawodowa). Każda z nich, usytuowana na innym jakościowo poziomie, stanowi przedmiot szczególnego rodzaju badań (w tym: fizycznych, biologicznych, psychologicznych, społecznych i kulturowych), które w odniesieniu do zachowań i przeżyć człowieka dostarczają innego rodzaju wyjaśnień. Por. (Scott 1999; Marciszewski, Stacewicz 2011).

W przypadku systemów informatycznych wielopoziomowość odgrywa równie istotną rolę, choć póki co, gdy nie istnieją jeszcze układy w pełni autonomiczne, jej analiza musi uwzględniać fakt, że systemy informatyczne są projektowane przez ludzi, do realizacji stawianych przez nich celów. Koncentrując uwagę *na aspektach programistycznych*, proponujemy wyróżnić, za pracą (Primiero 2016), **kilka powiązanych ze sobą poziomów, na których sytuują się różne jakościowo opisy procedur, prowadzących od danych wejściowych systemu do realizacji powierzonego systemowi zadania.** Poziomy te będziemy nazywać dalej *warstwami abstrakcji*⁶⁶, ponieważ na każdej kolejnej warstwie abstrahuje się od elementów decydujących o specyfice warstw poprzednich. Są to warstwy następujące: 1) warstwa *fizyczna*⁶⁷ – dotycząca sposobu przetwarzania sygnałów fizycznych, oddziaływujących na podłączone do komputera urządzenia fizyczne, 2) warstwa *kodu maszynowego* – uwzględniająca właściwości procesora kontrolującego przetwarzanie sygnałów fizycznych, w tym listę jego rozkazów, 3) właściwe *warstwy programistyczne* – dotyczące różnych języków programowania, w tym języków wysokiego poziomu), 4) warstwa *algorytmiczna* – ulokowana poza systemem, ale powiązana z nim przez fakt, że zarówno programy, jak i kody maszynowe są implementacjami różnych algorytmów, 5) warstwa *specyfikacji* – na której określa się rozwiązywany algorytmicznie problem, w tym warunki jego poprawnego rozwiązania.

Chociaż charakterystyki powyższe są sformułowane bardzo ogólnie, to w sposób oczywisty dotyczą one także tych fragmentów systemów przetwarzających informacje, które odpowiadają za operowanie pojęciami, w tym za podejmowanie konkretnych decyzji kategoryzacyjnych. Przykładowo: system rozpoznający stany gardła jako wskazujące (bądź nie) na infekcję wirusową, działa według pewnego programu (warstwa 3), program ten stanowi implementację konkretnego algorytmu (warstwa 4), algorytm zaś jest zgodny z taką a taką specyfikacją medyczną (warstwa 5).

⁶⁶ Jest to zgodne z terminologią przyjętą w pracy (Primiero 2016).

⁶⁷ Jej wyróżnienie w kontekście programistycznym jest o tyle wskazane, że komputery mogą być programowane czysto fizycznie przez odpowiednią konfigurację pewnych elementów fizycznych, takich jak przełączniki, połączenia między nimi, napięcia elektryczne (dotyczy to np. pierwszych maszyn cyfrowych, konstruowanych w latach 40. XX wieku).

Z uwagi na wielopoziomową strukturę systemów informatycznych – ale także modelowanych za ich pomocą systemów naturalnych – wynika arcyważne zagadnienie wyboru tej warstwy, która zawiera informacje najbardziej istotne z punktu widzenia wyjaśnień. Piszac „najbardziej istotne”, mamy na myśli informacje, które mogą być wykorzystywane również na innych warstwach, np. w celu adekwatnej interpretacji zawartych tam danych.

Stawiamy tezę, że **warstwą we wskazanym kontekście najważniejszą jest warstwa algorytmu**. Przemawiają za tym następujące argumenty. Po pierwsze, spośród wszystkich schematów, które określają sposób działania systemu (w tym: zbiorów instrukcji różnych języków programowania), algorytmy pozostają najbardziej czytelne. Zazwyczaj formułuje się je w języku naturalnym lub bardzo do niego zbliżonym, używając przy tym łatwo interpretowalnych elementów graficznych (jak grafy, etykietowane pola czy schematy blokowe). Tak przejrzysty sposób opisu ułatwia formułowanie wyjaśnień. Po drugie, w poprawnie zaprojektowanym systemie algorytm określa ściśle metodę rozwiązania problemu. Można ją wprowadzić zaimplementować za pomocą różnych języków programowania, ale wszelkie zgodne z danym algorytmem programy będą realizować tę samą metodę i gwarantować identyczne wyniki⁶⁸. Algorytm stanowi zatem nie tylko opis czytelny, ale również adekwatny. Po trzecie wreszcie, **warstwa algorytmu pozostaje najbliższa specyfikacji problemu**. Chociaż służy określeniu możliwej do zaprogramowania metody dochodzenia do wyników, to jest powiązana ściśle z opisem samego problemu i warunków poprawności rozwiązań. Z tego powodu, odwołując się do tej warstwy, najłatwiej jest uzyskać wyjaśnienia w języku, w jakim sformułowano sam problem. To zaś uznaliśmy wcześniej za istotny warunek przejrzystości poznawczej systemu.

Konkludując, możemy stwierdzić, że systemy operujące pojęciami (w tym: rozpoznające ich desygnaty), ze względu na wielość warstw pozwalających opisać ich sposób działania, zapewniają różne typy wyjaśnień. Przykładowo: wyjaśnienia istotne dla programistów różnią się od wyjaśnień oczekiwanych przez zwykłych użytkowników. Mimo to, warstwą uprzywilejowaną, która zapewnia wystarczająco klarowne wyjaśnienia na różnych poziomach (dostosowanych do różnego typu wymagań), jest warstwa algorytmu.

⁶⁸ Z wielością programistycznych implementacji algorytmu wiąże się pewien problem. Otóż znając konkretny program, nie zawsze potrafimy odtworzyć algorytm, któremu on odpowiada. W szczególności, jeśli program powstaje w drodze uczenia się (interakcji ze środowiskiem) przez dokonywanie pożądanych (skutecznych) zmian w kodzie samego programu, to w programie mogą pojawiać się pewne techniczne parametry czy ciągi kodowe, których nie będziemy potrafili przełożyć na zapis algorytmiczny. Nie będziemy zatem w stanie odpowiedzieć na pytanie, jakiego rodzaju algorytm jest realizowany. Por. (Creel 2020).

5.4. Podsumowanie

1. Zgodnie z metodologią **psychologii behawioralnej i cybernetyki** (dość popularną do lat 60. XX wieku) do modelowania niektórych zjawisk czy procesów (w tym: umysłowych) należy stosować **metodę czarnej skrzynki**, która polega na konstrukcji modelu wyłącznie na podstawie statystycznie istotnych **korelacji** między obserwowalnymi bodźcami (wejściem) i reakcjami (wyjściem). W świetle tej metody czarną skrzynką jest obiekt modelowany – obiekt o **nieprzejrzystej** strukturze wewnętrznej.

2. Współczesny problem czarnej skrzynki, zwany też problemem nieprzejrzystości poznawczej, dotyczy **struktury wyjaśnień**, a nie struktury relacji wiążącej wejście z wyjściem. Na polu informatyki „czarnymi skrzynkami” nazywa się układy, które generują poprawne wyniki, mają też pewną intersubiektywnie dostępną „zawartość”, nie zapewniają jednak (i nie przedstawiają) **zrozumiałych dla człowieka wyjaśnień** uzyskiwanych wyników.

3. Do najważniejszych **powodów nieprzejrzystości poznawczej** (w sensie eksplanacyjnym) należą: **złożoność** strukturalna, **interakcja** systemu ze środowiskiem (w tym: uczenie się), naśladowanie **naturalnych** układów do przetwarzania informacji, operowanie na danych **niepewnych**, brak procedur **translacji** operacyjnych reguł niskiego poziomu na czytelne dla człowieka reguły symboliczne.

4. Informatyczne reprezentacje pojęciowe są zróżnicowane ze względu na **stopień swojej przejrzystości**: najmniej przejrzyste są reprezentacje konekcyjne i genetyczne, najbardziej przejrzyste – reprezentacje logiczystyczne.

5. Ze względu na **wielopoziomową** strukturę systemów operujących pojęciami istnieją różne typy wyjaśnień, formułowane na różnych poziomach. W przypadku systemów informatycznych poziomem, który zapewnia wyjaśnienia najbardziej czytelne, jest poziom **algorytmów**.

Bibliografia

- Ashby W.R., *Design for a Brain*, 1st ed. John Wiley & Sons, New York/London 1952.
- Ajdukiewicz K., *Logika pragmatyczna*, PWN, Warszawa 1965.
- Ashby W.R., *Wstęp do cybernetyki*, PWN, Warszawa 1961.
- Bolc L., *Many-valued Logic*, Springer-Verlag, Berlin 1992.
- Bolc L., Cytowski J., Stacewicz P., *O logice i wnioskowaniu przybliżonym*, Prace IPI PAN, Warszawa 1996.
- Chisholm R.M., *Teoria poznania*, Inst. Wydawniczy „Daimonion”, Lublin 1994.
- Cichosz P., *Systemy uczące się*, Wydawnictwa Naukowo-Techniczne, Warszawa 2000.
- Creel, K.A. *Transparency in Complex Computational Systems*. Philosophy of Science 87.4, 2020, s. 568–589. doi: 10.1086/709729. url: <https://doi.org/10.1086/709729>.
- Chlewiński Z. (red), *Modele umysłu*, PWN, Warszawa 1999.
- Czogała E., Pedrycz W., *Elementy i metody teorii zbiorów rozmytych*, PWN, Warszawa 1985.
- Flasiński M., *Wstęp do sztucznej inteligencji*, Wydawnictwo Naukowe PWN, Warszawa 2011.
- Folger T.A., Klir G.J., *Fuzzy Sets, Uncertainty and Information*, Prentice Hall, Englewood Cliffs 1988.
- Frege G., *Pisma semantyczne*, PWN, Kraków 1977.
- Gärdenfors P., *Conceptual Spaces: The Geometry of Thought*, MIT Press, 2004.
- Hinton G.E., *Connectionist learning procedures*, Artificial Intelligence (40), 1989, s. 185-234.
- Hunt E.B., Marin J., Stone P.T., *Experiments in Induction*, Academic Press, New York 1966.
- Ignizio J.P., *An Introduction to Expert Systems*, McGraw-Hill, New York 1991.
- Kari L., Rozenberg G. *The many facets of natural computing*. Communications of the ACM, 10(51), 2008, s. 72–83.
- Klir G., Yuan B., *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River 1995.
- Kotarbiński T., *Elementy teorii poznania, logiki formalnej i metodologii nauk*, PWN, Warszawa 1986.
- Kuratowski K., Mostowski A., *Teoria mnogości*, PWN, Warszawa 1978.
- Leśniewski S., *Podstawy ogólnej teorii mnogości. I.*, [w:] *Prace Polskiego Koła Naukowego w Moskwie*, Sekcja matematyczno-przyrodnicza, no. 2, Moskwa 1916.
- Leśniewski S., *O podstawach matematyki*, Wstęp, Rozdziały I-XI, *Przegląd Filozoficzny*, Warszawa 1927-1931.
- Lindsay P.H., Norman D.A., *Procesy przetwarzania informacji u człowieka*, PWN, Warszawa 1984.
- Marciszewski W., *Podstawy logicznej teorii przekonań*, PWN, Warszawa 1972.
- Marciszewski W., Stacewicz P., *Umysł – Komputer – Świat. O zagadce umysłu z informatycznego punktu widzenia*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2011.

- Maruszewski T., *Psychologia poznawcza*, Biblioteka Myśli Semiotycznej, Warszawa 1996.
- Najder K., *Reprezentacje i ich reprezentacje*, Zakład Narodowy im. Ossolińskich, Wrocław 1989.
- Najder K., *Pojęcie w psychologii*, [w:] Kurcz I. (red.) *Psychologia a semiotyka*, Biblioteka Myśli Semiotycznej, Warszawa 1993.
- Pawlak Z., *Systemy informacyjne. Podstawy teoretyczne*, WNT, Warszawa 1983.
- Pawlak Z., *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht 1991.
- Pawłowski T., *Tworzenie pojęć i definiowanie w naukach humanistycznych*, PWN Warszawa 1978.
- Pelc J., *O użyciu wyrażen*, Ossolineum, Wrocław 1971.
- Piętka D., Stacewicz P., *A Decision Logic Approach to Mill's Eliminative Induction*, "Studies in Logic, Grammar and Rhetoric", 42 (55), 2015, s. 113-138.
- Piłat R., *O istocie i pojęć*, Wydawnictwo Instytutu Filozofii i Socjologii PAN, Warszawa 2007.
- Primiero G., *Information in the Philosophy of Computer Science*, in L. Floridi (ed.) *The Routledge Encyclopedia on the Philosophy of Information*, 2016, s. 90-106.
- Russel S., Norvig P. *Artificial Intelligence: A Modern Approach*, Pearson, London 2020.
- Ryle G., *Czym jest umysł*, PWN, Warszawa 1970.
- Scott A., *Schody do umysłu*, Wydawnictwa Naukowo-Techniczne, Warszawa 1999.
- Skinner B.F., *The Behavior of Organisms*, New York 1938.
- Skowron B., *Część i całość. W stronę topontologii*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2021.
- Smith E.E., Medin D.L., *Categories and Concepts*, Harvard University Press, Cambridge 1981.
- Stacewicz P., *Uogólniona reguła modus ponens. Metoda wnioskowania z danych rozmytych*, Prace IPI PAN, zeszyt nr 742, 1994.
- Stacewicz P., *Umysł a modele maszyn uczących się. Współczesne badania informatyczne w oczach filozofa*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2010.
- Stacewicz P., *O algorytmach i algorytmicznej dostępności wiedzy*, *Studia Metodologiczne*, 2016, nr 36, s. 315-331.
- Tadeusiewicz R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa 1993.
- Terano T., Asai K., Sugeno M., *Applied Fuzzy Systems*, Academic Press, London 1994.
- Tulving E., *Episodic and semantic memory*, [w:] Tulving E., Donaldson W. (red.), *Organization of memory*, Academic Press, New York 1972.
- Turing A.M., *On computable numbers, with an application to the Entscheidungsproblem*, *Proceedings of the London Mathematical Society* s2-42(1), 1936, s. 230-265.
- Turing A.M., *Computing Machinery and Intelligence*, *Mind* (59/236), 1950.
- Zadeh L.A., *Calculus of fuzzy restrictions*, [w:] Zadeh L.A. /i in/ (red.), *Fuzzy Sets and their applications to cognitive and decision processes*, Academic Press, New York 1975, s. 1-25.
- Zednik, C. *Solving the Black Box Problem: A Normative Framework for Explainable*, *Philosophy & Technology*, 2019, <https://doi.org/10.1007/s13347-019-00382-7>.
- Ziemiński Z., *Logika praktyczna*, wydanie XXIII PWN, Warszawa 2000.

Indeks rzeczowy

algorytm 12, 28, 44, 48nn, 56, 58n, 68nn
algorytmizacja 52nn, 58nn, 62nn

behawioryzm 61nn, 72

cybernetyka 54, 61nn, 72
czarna skrzynka 8, 60nn, 72

dane rozmyte 35n
decyzja kategoryzacyjna 7, 27nn, 48nn, 66nn
definiowanie 11, 17nn, 27, 33n, 47, 69
drzewa decyzyjne 19, 25, 53, 55m 59, 68

funkcje decyzyjne 7, 30, 33, 49n, 52
funkcje logiczne 24n
funkcje w matematyce 23

informatyczne reprezentacje pojęć 49n, 52nn,
58n, 67n, 72
informatyczne systemy uczące się 64, 66

kategoryzacja 27nn, 33, 50nn, 66nn
komunikacja człowieka z maszyną 37, 49,
61, 66
kontekst działania 57
kontekst wyjaśniania 58n

logiki rozmyte 37, 65
logiki wielowartościowe 29, 34, 38, 47,
56, 65

mereologia 45nn
metoda czarnej skrzynki
metody reprezentacji wiedzy 8, 42, 49nn,
54nn, 58n, 64n
modelowanie czynności poznawczych 8, 48,
54, 67n, 72

nazwy a pojęcia 10, 12, 14nn, 17nn, 22n, 26n,
30, 46n, 66
nazwy, typologia 15nn
nazwy, znaczenie 15n
neurony w mózgu 13, 19, 25, 27n
nieprzejrzystość poznawcza 8, 57, 59, 61,
63nn, 72
norma logiczna 12

obliczenia naturalne 65
optymalizacja reprezentacji pojęciowych 51n,
58

pamięć semantyczna 13
percepcja 14, 52
pojęcia a zbiory 7, 31nn
pojęcia wg Fregego 7n, 24nn
pojęcie a funkcja 7n, 23nn, 31nn, 47n
pojęcie a nazwa 10, 12, 14nn, 17nn, 22n, 26n,
30, 46n, 66
pojęcie nominalne 10, 12n, 17, 21n, 49
pojęcie ostre 16, 25, 28, 30, 33nn, 43, 50, 57
pojęcie pewne 38nn
pojęcie przybliżone 29, 38nn
pojęcie rozmyte 33nn
pojęcie zbiorowe 44nn
pojęcie, jako funkcja 7n, 23nn, 31nn, 47n,
49n, 59
pojęcie, reprezentacja 18nn, 66nn
pojęcie, zakres 7, 19nn, 24nn, 29nn, 31nn,
40n, 45nn, 58n
postulat ostrości pojęć 25, 28, 30
problem czarnej skrzynki 8, 28, 60nn, 72
procedury translacji 65, 67, 72
przejrzystość poznawcza 8, 55n, 58, 60nn,
66nn
przekonania 10nn, 21

- przybliżenie dolne zbioru 40n, 44, 47
przybliżenie górne zbioru 40n, 44, 47
- relacja nierozróżnialności 41nn
- reprezentacje koneksyjne 28, 49, 56nn, 69, 72
- reprezentacje logicystyczne 54, 56, 59, 72
- reprezentacje naturalistyczne 54, 56, 59
- reprezentacje pojęciowe 8, 13n, 18nn, 25nn, 48nn, 66nn
- reprezentacje pojęciowe, egzemplarzowe 20n, 66
- reprezentacje pojęciowe, klasyczne 19nn
- reprezentacje regułowe 49, 51, 55
- reprezentacje subsymboliczne 56n
- reprezentacje symboliczne 54nn, 62, 64n, 68, 72
- rozpoznawanie desygnatów pojęć 50
- rozumowanie 10, 14, 18, 52
- semiotyczne, własności 16nn
- semiotyka 14n, 17, 22, 35
- sieci semantyczne 13, 55, 68
- system informatyczny 42, 48nn, 53nn, 58n, 60nn
- systemy eksperckie 52, 54n, 58, 64
- sztuczna inteligencja 8n, 27, 37, 49nn, 62, 67
- sztuczne sieci neuronowe 8, 28, 53, 56nn, 62nn
- tworzenie pojęć 17, 51n, 53, 58n
- uczenie się pojęć 49nn, 58n, 64nn, 72
- uzasadnianie 11n, 21, 63, 69
- warstwy abstrakcji 70
- wiedza 10nn, 40nn, 49nn, 59, 62nn
- wiedza w sensie logicznym 10nn, 21
- wiedza w sensie psychologicznym 10nn, 21
- wyjaśnianie 52, 54n, 59, 67nn
- wyjaśnianie decyzji kategoryzacyjnych 52, 69nn
- zbiory przybliżone 29, 38nn
- zbiory rozmyte 29, 34nn
- zbiory, w sensie dystrybutywnym 31n
- zbiory, w sensie kolektywnym 31
- złożoność pojęć 69
- złożoność strukturalna 64
- znaczenie nazwy 10, 12nn, 20nn, 28

Indeks nazwisk

Ajdukiewicz K. 12, 14, 17n
Arystoteles 24
Asai K. 37
Ashby W.R. 9, 54, 61
Austin J.L. 15

Bolc L. 34, 44

Chisholm R.M. 11
Chlewiński Z. 18
Chomsky N. 54
Creel K.A. 64, 71
Cytowski J. 44
Czogała E. 35, 38, 49

Flasiński M. 27, 55
Frege G. 7n, 15, 24nn, 33, 48n, 59n

Ganter B. 49

Hinton G.E. 56, 65
Hunt E.B. 55

Ignizio J.P. 55

Jadacki J. 11

Kari L. 56
Klir G. 29, 34, 65
Kotarbiński T. 12, 14, 16
Kuratowski K. 45

Leśniewski S. 45, 47
Lindsay P.H. 13
Locke J. 15
Marciszewski W. 11, 70
Marin J. 55

Maruszewski T. 13n, 18n, 74
Medin D.L. 19, 36
Mill J.S. 15
Morris Ch. 15
Mostowski A. 45

Najder K. 19, 21, 33
Norman D.A. 13
Norvig P. 49

Omyła M. 9

Pawlak Z. 29, 39, 41nn, 55
Pedrycz W. 35, 38n
Pelc J. 14n
Piętka D. 42n
Primiero G. 70

Rozenberg G. 56
Russel S. 49
Ryle G. 61

Scott A. 70
Skowron B. 46, 64
Słowiński R. 42
Smith E.E. 18, 36
Stacewicz P. 36, 38, 49n, 54, 58
Stone P.T. 55
Strawson P.F. 54
Sugeno M. 54

Tadeusiewicz R. 56n
Terano T. 37
Tulving E. 13
Turing A.M. 9, 54, 62,

Wille R. 49

Wittgenstein L. 15

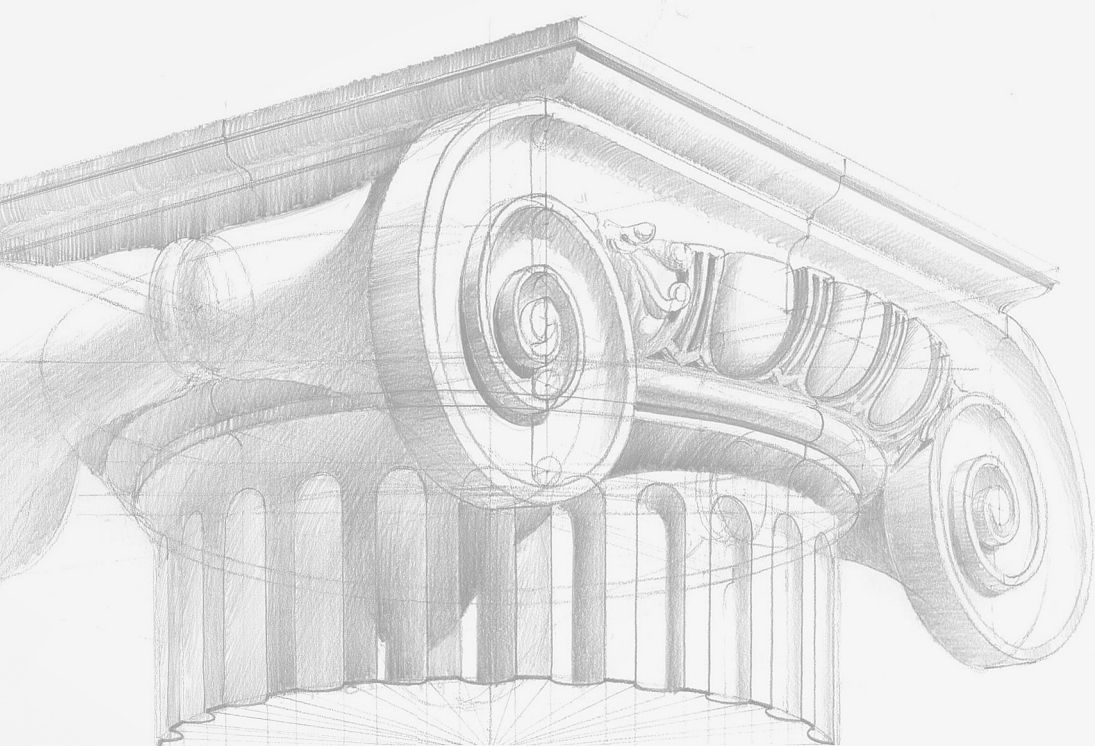
Włodarczyk A. 9

Yuan B. 29, 65

Zadeh L.A. 61

Zednik C. 62

Ziemiński Z. 63



Czym są pojęcia, że możemy je pojmować na różne sposoby, wnikając w różne warstwy wytwarzanej i gromadzonej przez człowieka wiedzy? Pytanie to nurtuje filozofów, psychologów, semiotyków, logików, lingwistów, ale także informatyków i specjalistów od sztucznej inteligencji. Wszyscy oni uznają pojęcia za podstawowy składnik wiedzy, który umożliwia systemom opartym na wiedzy – tak naturalnym, jak sztucznym – podejmowanie decyzji kategoryzacyjnych. Umożliwia zatem percepcję, zapamiętywanie i elementarne formy rozumowania.

Dlaczego pojęcia można rozumieć jako funkcje decyzyjne? Jak tego rodzaju funkcje są reprezentowane wewnątrz naturalnych i sztucznych systemów poznawczych? Dlaczego nie wszystkie ich reprezentacje są czytelne dla człowieka? W jaki sposób teorie pojęć odwołują się do teorii zbiorów? Są to najważniejsze zagadnienia, które podejmuje w swojej książce.

Paweł Stacewicz

ISBN 978-83-8156-275-1



9 788381 562751