

**Jadwiga Woźniak**

# **KATEGORYZACJA**

**WYDAWNICTWO**

**SBP**



**NAUKA-DYDAKTYKA-PRAKTYKA**

# **KATEGORYZACJA**

**Polish Librarians Association**  
**SCIENCE-DIDACTICS-PRACTICE**

**Jadwiga Woźniak**

**CATEGORISATION**

**A study in the theory  
of indexing languages**

**WYDAWNICTWO**

**SBP**



**Warsaw 2000**

**Stowarzyszenie Bibliotekarzy Polskich**  
**NAUKA-DYDAKTYKA-PRAKTYKA**

**Jadwiga Woźniak**

**KATEGORYZACJA**

**Studium z teorii języków  
informacyjno-wyszukiwawczych**

**WYDAWNICTWO  
SBP**



**Warszawa 2000**

Komitet Redakcyjny serii wydawniczej  
<<NAUKA — DYDAKTYKA — PRAKTYKA>>

Marcin DRZEWIECKI (przewodniczący), Stanisław CZAJKA, Zofia GACA-  
-DĄBROWSKA, Barbara SOSIŃSKA-KALATA, Danuta KONIECZNA,  
Krzysztof MIGOŃ, Mieczysław MURASZKIEWICZ, Janusz NOWICKI (sekretarz),  
Wanda PINDŁOWA, Jan SÓJKA, Barbara STEFANIAK, Hanna TADEUSIEWICZ,  
Zbigniew ŻMIGRODZKI

Książka została wydana przy wsparciu finansowym Instytutu Informacji Naukowej  
i Studiów Bibliologicznych UW

Recenzenci

prof. dr hab. Bożenna BOJAR  
prof. dr hab. Marcin DRZEWIECKI

Redakcja techniczna i korekta  
Anna LIS

© Copyright by Stowarzyszenie Bibliotekarzy Polskich

ISBN 83-87629-32-4

CIP — Biblioteka Narodowa

Woźniak, Jadwiga

Kategoryzacja : studium z teorii języków informacyjno-wyszukiwawczych / Jadwiga  
Woźniak ; Stowarzyszenie Bibliotekarzy Polskich. - Warszawa : Wydaw. SBP, 2000. -  
(Nauka, Dydaktyka, Praktyka ; 36)

Wydawnictwo SBP. Warszawa 2000. Wydanie I. Ark. wyd. 14.0. Ark. druk. 14.0.  
Łamanie: K. Brawiński. Druk i oprawa: Zakład „PRIMUM” Kozerki 17a,  
05-825 Grodzisk Maz.

## SPIS TREŚCI

|   |     |
|---|-----|
| <b>WPROWADZENIE</b> .....   | 7   |
| Rozdział pierwszy   |     |
| <b>JĘZYK A REPREZENTACJA DOKUMENTU W SYSTEMIE<br/>INFORMACYJNO-WYSZUKIWAWCZYM</b> ..... | 19  |
| Rozdział drugi  |     |
| <b>KATEGORYZACJA: SELEKTYWNY PRZEGLĄD BADAŃ</b> .....                                   | 65  |
| Kategorie Arystotelesowskie .....   | 71  |
| Kategoryzacja w semantyce językoznawczej .....  | 77  |
| Wittgensteina koncepcja podobieństwa rodzinnego .....                                   | 95  |
| Kategorie prototypowe .....   | 97  |
| Rozdział trzeci   |     |
| <b>KATEGORYZACJA W TEORII I PRAKTYCE JĘZYKÓW<br/>INFORMACYJNO-WYSZUKIWAWCZYCH</b> ..... | 117 |
| Kategorie w językach informacyjno-wyszukiwawczych .....                                 | 123 |
| Kategorie w Uniwersalnej Klasyfikacji Dziesiętnej .....                                 | 144 |
| Kategorie w językach hasel przedmiotowych .....   | 152 |
| Kategorie w Tezaurusie technologii żywności i przemysłu spożywczego ..                  | 162 |
| Kategorie w wyszukiwarkach internetowych .....  | 168 |
| Tworzenie kategorii w językach informacyjno-wyszukiwawczych .....                       | 174 |
| Kategoryzowanie informacji .....  | 188 |
| <b>WNIOSKI</b> .....  | 196 |
| <b>BIBLIOGRAFIA ZAŁĄCZNIKOWA</b> .....  | 201 |

# CONTENTS

|   |     |
|---|-----|
| <b>INTRODUCTION</b> .....   | 7   |
| <b>1. LANGUAGE AND DOCUMENT REPRESENTATION<br/>IN INFORMATION RETRIEVAL SYSTEMS</b> ..... | 19  |
| <b>2. CATEGORISATION: A SELECTICE REVIEW OF RESEARCH</b> .....                            | 65  |
| Aristotelian categories .....   | 71  |
| Categorisation in linguistic semantics .....  | 77  |
| Wittgenstein's concept of family resemblance .....  | 95  |
| Prototype categories .....  | 97  |
| <b>3. CATEGORISATION IN THE THEORY AND PRACTICE<br/>OF INDEXING LANGUAGES</b> .....       | 117 |
| Categories in indexing languages .....  | 123 |
| Categories in the Universal Decimal Classification .....                                  | 144 |
| Categories in subject headings systems .....  | 152 |
| Categories in the Thesaurus of Food Technology and Food Industry .....                    | 162 |
| Categories in the Internet search engines .....   | 168 |
| The development of category-based indexing languages .....                                | 174 |
| Categorisation of objects .....   | 188 |
| <b>CONCLUSION</b> .....   | 196 |
| <b>BIBLIOGRAPHY</b> .....   | 201 |

## WPROWADZENIE

Głównym przedmiotem rozprawy jest kategoryzacja leksykalna w językach informacyjno-wyszukiwawczych (jiw), stosowanych w niepełnotekstowych<sup>1</sup> dokumentacyjnych systemach wyszukiwania informacji (siw).

Istotą kategoryzacji jest postrzeganie podobieństw w różnorodności, co implikuje ważne problemy filozoficzne, w tym pytania o naturę znaku językowego. Na jednym biegunie rozważań o naturze znaków językowych lokuje się postawa nominacyjna<sup>2</sup>, zakładająca że podobieństwo jest jedynie kwestią konwencji lingwistycznej – istoty żywe, które można nazwać psami czy zbiór kolorów, które mogą być określone jako czerwony, w rzeczywistości nie mają nic wspólnego poza nazwą. Na drugim biegunie lokuje się realizm<sup>3</sup>, głoszący że kategorie takie jak PIES czy CZERWONY istnieją niezależnie od języka i jego użytkowników oraz że słowa *pies* i *czerwony* tylko nazywają te istniejące, nie tworzone, kategorie. Jeszcze inaczej próbuje objaśniać ten problem konceptualizm<sup>4</sup>, w którym zakłada się, że pomiędzy znakami języka i bytami, do których znaki się odnoszą, istnieje byt mentalny, tj. pojęcie. Efektem znajomości pojęć ‘pies’ czy ‘czerwony’, inaczej mówiąc, efektem wiedzy mówiącego o znaczeniu wyrazów *pies* i *czerwony*, jest kategoryzacja pewnych istot jako psy, zaś pewnych kolorów jako czerwone itd.

W teorii języków informacyjno-wyszukiwawczych problem znaczenia znaków językowych ma dodatkowy aspekt. Język informacyjno-wyszukiwawczy, zwany też językiem informacyjnym czy językiem indeksowania, jest językiem sztucznym przeznaczonym do pełnienia dwóch zasadniczych funkcji – metainformacyjnej, wyrażającej się w tworzeniu reprezentacji dokumentów (charakterystyk wyszukiwawczych) i pytań użytkowników (instrukcji wyszukiwawczych), oraz wyszukiwawczej, polegającej na inicjowaniu i kształtowaniu procesu wyszukiwania informacji. W celu zapewnienia realizacji tych funkcji projektanci jiw dążą do normalizacji planów treści i wyrażania języków informacyjno-wyszukiwawczych między innymi poprzez tworzenie systemów notacyjnych czy poprzez znormalizowane i kontrolowane stosowanie znaków języka naturalnego (zjawisko quasi-naturalności słownictwa pewnych jiw). Do tej pory w odniesieniu do języków informacyjno-wyszukiwawczych stosowano przede wszystkim dwie koncepcje znaczenia: rozumienie znaczenia jako denotacji wyrażań oraz rozumienie znaczenia jako konotacji wyrażań. W rozprawie podejmuję próbę upraw-

---

<sup>1</sup> Wiele z zawartych w pracy stwierdzeń jest prawdziwych także w odniesieniu do systemów pełnotekstowych.

<sup>2</sup> Językowa 1996; Kotarbiński 1957; Lakoff 1987; Taylor 1991.

<sup>3</sup> Ibidem.

<sup>4</sup> Ibidem.



dopodobnienia użyteczności jeszcze jednego podejścia – rozumienia znaczenia jako użycia znaku w systemie operującym reprezentacjami metatekstowymi. Ujmowanie znaczenia jako użycia może i powinno, moim zdaniem, mieć zastosowanie w projektowaniu jiw jako metoda komplementarna, dopełniająca podejście denotacyjne i konotacyjne.

W literaturze terminy *system dokumentacyjny* – *system informacji dokumentacyjnej* – *system informacji pośredniej* – *system pośredni* – *system skierowujący* oraz *system faktograficzny* – *system informacji faktograficznej* – *system informacji bezpośredniej* – *system bezpośredni* są traktowane jako dwa odrębne łańcuchy synonimów. W rozprawie częściej posługuję się terminami *system dokumentacyjny* i *system faktograficzny* ze względu na ich krótkość i większe rozpowszechnienie, choć bardziej jednoznaczne wydają się być terminy wskazujące na bezpośredni lub pośredni charakter odpowiedzi systemu, jako że przy pewnym rozumieniu informacji dokumentacyjnej (jako określonego podzbioru informacji faktograficznej) wszystkie systemy informacyjno-wyszukiwawcze mogą być uważane za systemy faktograficzne.

Jednoznaczne ustalenie cech dystynktywnych obu typów systemów nie jest łatwe. *Słownik encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych* definiuje te dwa rodzaje systemów, nazywając je odpowiednio systemami informacji dokumentacyjnej i systemami informacji faktograficznej. Z systemem informacji dokumentacyjnej mamy do czynienia wtedy, „gdy na podstawie informacji wejściowej tworzy się zbiór metainformacji w postaci charakterystyk wyszukiwawczych dokumentów poprzez transformacje streszczania, wyboru słów kluczowych i indeksowanie przy użyciu odpowiedniego (używanego w tym systemie) języka informacyjno-wyszukiwawczego, przy czym streszczanie jest operacją fakultatywną”<sup>5</sup>. O systemie informacji faktograficznej mówimy wówczas, „gdy na wyjściu system udostępnia odpowiednio wyselekcjonowane informacje wejściowe, w wypadku tzw. systemów inteligentnych poddane odpowiednim transformacjom”<sup>6</sup>.

W literaturze spotykane jest przeciwstawianie systemu dokumentacyjnego systemowi pełnotekstowemu, jakby to były klasy współrządne wyodrębnione w wyniku podziału systemów informacyjno-wyszukiwawczych według jednego kryterium. Opozycja taka jest fałszywa i nie wnosi niczego istotnego ani do teorii, ani do praktyki systemów i języków informacyjno-wyszukiwawczych. Wprawdzie system dokumentacyjny nie jest gatunkiem systemu pełnotekstowego, ale systemy pełnotekstowe z pewnego punktu widzenia mogą być postrzegane jako podzbiór zbioru systemów dokumentacyjnych.

Ważne z punktu widzenia pracy różnice między systemem dokumentacyjnym a systemem faktograficznym można wyrazić następująco. System dokumentacyjny nie wyszukuje informacji żądanych przez użytkownika, a jedynie przekazuje mu dane, najczęściej opisy bibliograficzne dokumentów lub ich teksty (system pełnotekstowy), w których prawdopodobnie będą się znajdować potrzebne użytkownikowi informacje. Inaczej mówiąc, system dokumentacyjny udziela użytkownikowi odpowiedzi pośredniej, skierowuje go do potencjalnie użytecznych tekstów (dokumentów), wykorzystując między innymi zbiór ich metainformacyjnych reprezentacji. Pełny tekst dokumentu

---

<sup>5</sup> Słownik 1993a, s. 148.

<sup>6</sup> Ibidem.

rozcłonkowany na poszczególne wyrażenia, na przykład z wyjątkiem tych, które pokrywają się z elementami tzw. stop listy, jest także jego metainformacyjną reprezentacją, choć inaczej utworzoną i o innym charakterze.

Systemy faktograficzne z logicznego punktu widzenia są deterministyczne, podczas gdy systemy dokumentacyjne są probabilistyczne<sup>7</sup> w sensie związku między zapytaniem informacyjnym a szansą, że użytkownik będzie usatysfakcjonowany udzieloną odpowiedzią. Jeśli to samo zapytanie skierujemy do dwóch systemów faktograficznych operujących tym samym zasobem informacji, to otrzymamy taką samą odpowiedź. Jeśli natomiast identyczne zapytanie skierujemy do dwóch systemów dokumentacyjnych operujących na identycznym zbiorze dokumentów prymarnych<sup>8</sup>, to może się zdarzyć, że uzyskamy różne odpowiedzi. Przyczyną takiej sytuacji jest skomplikowana natura procesu tworzenia i manipulowania reprezentacjami dokumentów oraz różnice w stosowanych językach informacyjno-wyszukiwawczych. W odniesieniu do systemów faktograficznych sensowne jest pytanie o to, czy system odpowiedział poprawnie na zadane pytanie. Systemy dokumentacyjne można oceniać jedynie w kategoriach użyteczności, kryterium bardziej subiektywnego niż poprawność i, co za tym idzie, trudniej poddającego się pomiarowi.

Do wymienionych już różnic między systemem faktograficznym a systemem dokumentacyjnym dodać należy jeszcze jedną, wyrażającą się w związku między wyszukanyimi informacjami a pytaniem, w odpowiedzi na które informacje te zostały wyszukane. W przypadku systemu faktograficznego pytający może tylko w jeden lub na niewiele sposobów określić, jakiej informacji poszukuje. W przypadku systemu dokumentacyjnego ten sam dokument, traktowany analogicznie jako jednostkowa porcja informacji, może być reprezentowany i szukany na wiele różnych sposobów, na przykład poprzez wskazanie autora, tytułu, słów kluczowych w tytule, daty publikacji, różnych haseł przedmiotowych, symboli różnych klasyfikacji itd. oraz kombinacji tych cech. Mechanizm wyszukiwania informacji polega na porównywaniu zapytania użytkownika z pewnymi elementami tekstów dokumentów lub ich reprezentacji. Efekty wyszukiwania w systemie dokumentacyjnym w dużym, jeśli nie największym, stopniu zależą zatem od jakości reprezentacji dokumentów i od samego języka informacyjno-wyszukiwawczego. Choć zgoda co do prawdziwości powyższego sformułowania jest powszechna zarówno wśród teoretyków jak i praktyków informacji, to jego implikacje,

---

<sup>7</sup> Probabilistyczny charakter reprezentacji dokumentów przez lata był uważany za czynnik o charakterze marginalnym, wynikający z niestaranności czy nieracjonalności indeksatorów, a co za tym idzie możliwy do wyeliminowania z procesów informacyjno-wyszukiwawczych.

<sup>8</sup> Dokument prymarny (pierwotny, źródłowy) – termin trudny do zdefiniowania i objaśnienia, gdyż cecha prymarności jest komplementarna z cechą pochodności (dokument pochodny). Inaczej mówiąc, dokument wtedy jest prymarny, kiedy nie jest pochodny. „Dokument pierwotny, dokument prymarny, dokument źródłowy – dokument stanowiący podstawę do sporządzenia dokumentu pochodnego. Dokument pochodny – dokument zawierający metainformację o innym dokumencie (dokumentach), na przykład opis formalny dokumentu (dokumentów), i/lub opis rzeczowy (tego) dokumentu (dokumentów)”. Słownik 1993a, s. 35. „Dokumenty pierwotne (prymarne), publikacje przedstawiające informacje o wynikach badań naukowych (lub o faktach z zakresu życia praktycznego) w formie oryginalnej, nieskróconej, zgodnej z ustaleniami autora. D. pierwotne stanowią podstawę opracowania dokumentów pochodnych. Dokumenty pochodne, materiały informacyjne opracowane na podstawie dokumentów pierwotnych, jak bibliogr., tematyczne zestawienia dokumentacyjne, przeglądy piśmiennictwa, opracowania syntetyczne omawiające postępy wiedzy w poszczególnych dziedzinach, karty dokumentacyjne itp.” Encyklopedia 1976, s. 100.

zwłaszcza praktyczne, nie zawsze są już tak oczywiste i akceptowane. Warto w tym miejscu przypomnieć, że w tzw. prostych systemach faktograficznych (niededukcyjnych) typu obiekt-cecha środkiem służącym do identyfikacji obiektów, ich cech i wartości często jest język deskryptorowy ze słownikiem ustrukturalizowanym podobnie do tezauryusa oraz rozbudowanymi wykładnikami relacji syntagmatycznych<sup>9</sup>.

Teoria jiw jest istotną i „mocną” częścią dorobku nauki o informacji. Niektórymi zagadnieniami, na przykład znaczeniem form językowych jiw, strukturą języka czy problemami gramatyki jiw, zajmowali się przede wszystkim (jeśli nie jedynie) badacze polscy głównie związani z prowadzonym w Katedrze Lingwistyki Formalnej Uniwersytetu Warszawskiego najpierw przez profesora dr. Olgierda Adriana Wojtasiewicza, a następnie przez profesor dr hab. Bożennę Bojar, seminarium na temat teoretycznych podstaw języków informacyjno-wyszukiwawczych<sup>10</sup>. Nadal jest jednak wiele obszarów problemowych, które albo nie były przedmiotem badań, albo nie zostały zbadane wyczerpująco czy satysfakcjonująco. Postęp nauki i zmiany w środowisku funkcjonowania języków informacyjno-wyszukiwawczych sprawiają też, że pewne zagadnienia wymagają rewizji czy ponownego zbadania.

W rozprawie podejmuję próbę podsumowania zarejestrowanej w publikacjach wiedzy o językach informacyjno-wyszukiwawczych w zakresie określonym przez przedmiot główny rozprawy, którym jest kategoryzacja. Będę też dążyć do uzupełnienia wiedzy tam, gdzie jest to możliwe na obecnym etapie rozwoju teorii jiw i dyscyplin, z których dorobku teoria jiw i nauka o informacji czerpią zasilenia metodologiczne i informacyjne. Nie jest moim celem stworzenie konkretnej kategoryzacji czy określenie wszystkich możliwych implikacji przeniesienia dorobku różnych dziedzin wiedzy (głównie językoznawstwa i psycholingwistyki) na grunt nauki o informacji. Uwagę skoncentruję na wielostronnym i wnikliwym scharakteryzowaniu dostrzeżonego problemu, traktując przeprowadzone badanie jako pierwszy i podstawowy etap wszelkich innych działań. Można zatem powiedzieć, że praca stanowi wprowadzenie do obszernej problematyki kategoryzacji w jiw, porządkując dotychczasowy stan wiedzy, uzasadniając potrzebę, możliwości i wskazując kierunki dalszych badań oraz że staram się w niej ukształtować pewien, trochę odmienny od dotychczasowego, sposób postrzegania języków i systemów informacyjno-wyszukiwawczych.

Osnowę rozprawy, na której zbudowana została jej treść, stanowi siedem następujących tez:

1. Na efektywność dokumentacyjnego systemu informacyjno-wyszukiwawczego istotny wpływ ma między innymi jakość reprezentacji dokumentów i pytań użytkowników, będąca z kolei pochodną jakości języka informacyjno-wyszukiwawczego oraz pragmatyki jego stosowania w systemie, w tym jakości pracy osób tworzących charakterystyki (także instrukcje) wyszukiwawcze, oraz znajomość społecznego środowiska funkcjonowania systemu. Skuteczność funkcjonowania języka informacyjno-wyszukiwawczego nie jest bowiem pochodną tylko i wyłącznie wewnątrzjęzykowych cech samego jiw; zależy również m.in. od prawidłowości doboru języka do warunków konkretnego siw i oczekiwań jego użytkowników.

---

<sup>9</sup> Biclicka 1984.

<sup>10</sup> Kurck-Kokocińska 1994.

2. Optymalizacja jiw może się dokonywać między innymi poprzez interdyscyplinaryzację badań oraz jawne określanie założeń metateoretycznych i przyjmowanych teorii, do których badający się odwołując, w tym poprzez powiązanie teorii i praktyki jiw z dorobkiem teoretycznym semantyki<sup>11</sup> językoznawczej.

3. Jednym ze sposobów optymalizacji informacyjnej (metainformacyjnej) i wyszukiwawczej funkcji jiw jest semantyczna strukturalizacja systemu leksykalnego języka, a zwłaszcza podejście kategoryzacyjne, pozwalające na ujawnianie *explicite* w języku tych elementów znaczeń, które wiążą znak języka z relewantnymi dla użytkownika elementami rzeczywistości pozajęzykowej.

4. Podejście kategoryzacyjne stwarza możliwość zbliżenia struktur jiw do struktury informacyjnej świata dokumentów i struktur wiedzy (użytkowników) związanych z doświadczaniem rzeczywistości; sztywność właściwa tradycyjnym schematom klasyfikacyjnym (najczęściej nawiązującym do klasyfikacji nauk), będąc gwarantem stabilności jiw, jest jednocześnie przeszkodą w dynamicznym reagowaniu na zmiany i rozwój, także wiedzy użytkownika.

5. Kategorie w jiw mogą być albo kategoriami klasycznymi, albo kognitywnymi (w tym opartymi na prototypie); wybór powinien zależeć nie od apriorycznie przyjętych założeń a od właściwości elementów kategoryzowanego uniwersum.

6. Fałszywe wydaje się być przekonanie o istnieniu tylko jednej, optymalnej charakterystyki wyszukiwawczej dokumentu, aczkolwiek w praktyce biblioteczno-informacyjnej z różnych powodów, głównie finansowych, pewne reprezentacje dokumentów mogą być uznane za „jedyne i optymalne”.

7. Wykorzystanie podejścia kategoryzacyjnego wymaga wcześniejszego uporządkowania jego podstaw teoretycznych, w tym terminologii, i dotychczasowego dorobku praktycznego.

Rozprawa składa się z trzech rozdziałów. Pierwszy poświęcony jest reprezentacji informacji zawartych w dokumentach i informacji o samych dokumentach. W rozdziale tym przedstawiam takie zagadnienia jak: pojęcie przedmiotu dokumentu; definicje i sposoby rozumienia znaczenia w odniesieniu do znaków jiw; rolę definicji, eksplikacji i przykładów w uczeniu się i posługiwaniu znaczeniami w językach informacyjno-wyszukiwawczych; problem adekwatności reprezentacji dokumentów wobec samych dokumentów i oczekiwań użytkowników systemu; wykazując potrzebę badań z zakresu pragmatyki języków informacyjno-wyszukiwawczych oraz ściślejszego powiązania projektowania i używania jiw ze strukturą informacyjną świata dokumentów. Istotnym elementem rozważań jest analiza roli i wagi informacji kontekstowej w projektowaniu i używaniu języków informacyjno-wyszukiwawczych. W rozdziale znalazła się także informacja o zakresie i specyfice badań kognitywnych w nauce o informacji. Problematyka ta (między innymi ze względu na brak publikacji w języku polskim<sup>12</sup>) nie jest szerzej znana, dlatego zdecydowałam się poświęcić jej fragment rozdziału, aby wzbogacić tło, na którym dalej charakteryzuję i uzasadniam potrzebę optymalizacji podejścia do projektowania i stosowania języków informacyjno-wyszukiwawczych. Rozdział kończy opis badań prywatnych systemów informacji (księgozbiorów

---

<sup>11</sup> Wielu teoretyków nauki o informacji zajmowało się wcześniej tymi zagadnieniami (np.: Blair 1990; Foskett 1977; Sosińska 1988; Ungurian 1983, 1982, 1976; Vickery 1987a). Nie wszyscy jednak i nie zawsze łączyli swoje badania z teoriami powstałymi w sferze semantyki; szczególnie widać to w pracach autorów obcych.

<sup>12</sup> Woźniak 1997.

domowych, indeksów, kartotek itp.) historyków amerykańskich jako przykładów zastosowania w praktyce kategoryzowania i wyszukiwania informacji zasad intuicyjnego podejścia kognitywnego.

Rozdział drugi jest kontynuacją próby stworzenia tła przedmiotowego i metodologicznego dla kategoryzacji w językach informacyjno-wyszukiwawczych. Nawiązuję w nim do dorobku innych dyscyplin naukowych, w których kategoryzacja stanowi ważny problem badawczy. Zaczynam od rozważań terminologicznych na temat *kategoryzacji*, *kategoryzowania* i *kategorii*. Następnie podejmuję próbę przedstawienia wybranych koncepcji kategoryzacji, zaczynając od poglądów Arystotelesa jako prapoczątku wszelkich badań nad kategoryzacją, poprzez ujęcie kategoryzacji w strukturalistycznej semantyce językoznawczej, koncepcję podobieństwa rodzinnego wyrazów Wittgensteina aż do kognitywnych kategorii prototypowych. Przedmiotem rozprawy nie jest jednak zagadnienie kategoryzacji w ogóle, kategoryzacji językowej czy nawet kategoryzacji leksykalnej, a jedynie kategoryzacji leksykalnej w językach informacyjno-wyszukiwawczych. Dlatego rozdział drugi ma charakter selektywny, przeglądowy i w pewnym stopniu upraszczający. Mając tego świadomość, jak również pamiętając o interdyscyplinarnym charakterze problematyki kategoryzacji (cytowanie konkretnych publikacji jako prac z dziedziny filozofii, antropologii, psychologii czy językoznawstwa jest nierzadko sprawą umowną), chciałam przedstawić i przybliżyć przede wszystkim tło językowe i językoznawcze, na którym w rozdziale trzecim poddam analizie zjawisko kategoryzacji i jej realizacji w wybranych językach informacyjno-wyszukiwawczych, dla których relewantny i efektywny system kategorii ma ogromne znaczenie.

Rozdział trzeci skupia uwagę na problemach kategoryzacji w językach informacyjno-wyszukiwawczych. Trzy części rozdziału zawierają rozważania odpowiadające trzem sposobom rozumienia *kategoryzacji* w literaturze z zakresu języków i systemów informacyjno-wyszukiwawczych, tj. jako: systemu kategorii, procesu jego tworzenia oraz przypisywania obiektów rzeczywistości dokumentacyjnej i pozadokumentacyjnej do kategorii już istniejących w jiw.

Charakteryzując kategoryzację rozumianą jako system kategorii (*Kategorie w językach informacyjno-wyszukiwawczych*), zawęziłam pole badania do, o czym była już mowa, kategorii leksykalnych jiw. Za źródło materiału językowego przyjąłam cztery języki informacyjno-wyszukiwawcze – Uniwersalną Klasyfikację Dziesiątą, model języka haseł przedmiotowych stworzony przez Adama Łysakowskiego, język haseł przedmiotowych KABA oraz język deskryptorowy z dziedziny technologii żywności i przemysłu spożywczego, a właściwie jego słownik, tj. *Tezaurus technologii żywności i przemysłu spożywczego*. Badaniem objęłam również system kategorii dostępnych w wyszukiwarce internetowej AltaVista (stan na dzień 14 grudnia 1998). Wprawdzie nie jest w pełni jasne, jak są tu tworzone i używane kategorie, mimo to uznałam za celowe i ważne objęcie ich badaniem, zakładając, że (być może bez intencji jego autorów) system kategorii AltaVisty wraz z regułami reprezentowania i wyszukiwania dokumentów HTML może być postrzegany jako specyficzny rodzaj języka informacyjno-wyszukiwawczego w coraz liczniejszych i powszechniej wykorzystywanych zasobach informacyjnych Internetu. Zrezygnowałam z charakteryzowania większej liczby języków informacyjno-wyszukiwawczych, bo bez względu na to, jak bardzo są interesujące czy pouczające, pozostają tylko egzemplifikacją ogólnych stwierdzeń i postulatów, które są przedstawione w pracy. Moja uwaga koncentrowała się głównie na ka-

tegoriach leksykalnych. Wiele interesującego i wartościowego materiału na temat kategorii gramatycznych i syntaktycznych w jiw, choć przedstawionych w innym ujęciu, można znaleźć w niepublikowanej pracy doktorskiej Ewy Chmielewskiej-Gorczyca<sup>13</sup>.

Badając problemy związane z konstruowaniem systemu kategorii strukturalizujących pole semantyczne jiw (*Tworzenie kategorii w językach informacyjno-wyszukiwawczych*), korzystałam z dorobku językoznawstwa, w tym teorii pól semantycznych i metody ciągów definicyjnych, teorii języków informacyjno-wyszukiwawczych oraz koncepcji tzw. ontologii wywodzących się ze sztucznej inteligencji i inżynierii wiedzy. Autorzy publikacji z zakresu teorii i praktyki jiw rzadko podejmują problematykę dochodzenia do systemu kategorii i struktury semantyczno-leksykalnej języka informacyjno-wyszukiwawczego, co pozwala przypuszczać, że projektowanie struktury kategoryalnej jiw w dużej mierze ma charakter intuicyjny lub schematyczny. Wsparciem metodologicznym i przedmiotowym może tu służyć językoznawstwo i leksykologia.

Pojmowanie kategoryzacji jako procesu przyporządkowywania nowych obiektów do kategorii istniejących w jiw (*Kategoryzowanie informacji*) zostało w pracy scharakteryzowane poprzez wskazanie ważniejszych problemów badawczych i praktycznych lokujących się w tym obszarze pragmatyki jiw oraz ich implikacji. Znalezienie rozwiązań tych problemów wymaga odrębnych badań, których celowość i kierunki staram się uzasadnić. W rozdziale podejmuję też próbę określenia związku znaczeniowego między terminami indeksowanie i kategoryzowanie oraz wskazania ważniejszych problemów kategoryzacji tekstów literackich.

Pracę zamykają wnioski wynikające z przeprowadzonych badań. Badania były prowadzone metodą analizy krytycznej piśmiennictwa oraz obserwacji i analizy rzeczywistości funkcjonowania języków i systemów informacyjno-wyszukiwawczych. Jak wcześniej powiedziano, można je traktować jako pierwszy etap potencjalnego większego zamierzenia badawczego. Dlatego wśród wniosków są stwierdzenia odnoszące się do kierunków ewentualnych dalszych badań kategoryzacji w teorii i praktyce języków informacyjno-wyszukiwawczych.

*Bibliografia załącznikowa* rejestruje ponad 250 pozycji. Nie jest to kompletna bibliografia przedmiotu, a jedynie wybór wykorzystanych prac, z których przeważająca część jest cytowana w tekście rozprawy. Biorąc pod uwagę obecnie jedynie zalecający charakter norm bibliograficznych, po wnikliwym zapoznaniu się z normą PN-79/N-01222 *Kompozycja wydawnicza książki. Bibliografia załącznikowa* zdecydowałam się na opracowanie towarzyszącej rozprawie *Bibliografii* według reguł International Standard Bibliographic Description (ISBD).

*Bibliografia* powstała głównie w wyniku przeszukiwania drukowanych i elektronicznych wersji bibliografii *Library and Information Science Abstracts*, *MLA*<sup>14</sup> *International Bibliography*, *Psychological Abstracts*, *Przewodnik Bibliograficzny* oraz zawartości polskich czasopism naukowych i fachowych z zakresu bibliotekoznawstwa i nauki o informacji ze szczególnym uwzględnieniem „Przeglądu Bibliotecznego” i „Zagadnień Informacji Naukowej”. Nie można nie wspomnieć, że pewna część literatury, niewielka choć bardzo ważna, została zidentyfikowana nieformalnymi kanałami informacyjnymi, dzięki zainteresowaniu i pomocy osób, wśród których byli profesor Ia C. McIlwaine, dyrektor School of Library, Archive and Information Studies University

<sup>13</sup> Chmielewska-Gorczyca 1981.

<sup>14</sup> MLA – Modern Language Association of America.

College London oraz profesor Stephen Robertson i dr David Nicholas z City University Department of Information Science. Pozycje te byłyby trudne do odnalezienia w bibliografiach, katalogach czy innych bazach danych ze względu na przyjęte na ogół w tego typu zbiorach informacyjnych poziomy szczegółowości indeksowania, zalecające indeksować na przykład materiały z konferencji na poziomie ogólnej tematyki konferencji a nie tematyki poszczególnych wystąpień, czy kłopoty z identyfikacją relewantnych znaczeń w jiw stosowanych w tych bazach. Można powiedzieć, że zbierając materiały do *Bibliografii*, na sobie doświadczyłam wagi i znaczenia dobrego systemu kategoryzacji w języku informacyjno-wyszukiwawczym. Gromadzeniu literatury sprzyjały pobyty naukowe w zagranicznych uniwersytetach i bibliotekach, gdzie miałam okazję zapoznania się z materiałami niepublikowanymi, nieprzeznaczonymi do szerokiego rozpowszechniania i często nierejestrowanymi w spisach bibliograficznych. Szczególnie cenne okazały się zbiory Bibliotek Library Association, City University Department of Information Science w Londynie, Uniwersytetu Notre Dame w Indianie oraz Kent State University w Ohio (USA).

Największe problemy napotkałam przy gromadzeniu literatury do rozdziału trzeciego. Proces dochodzenia do systemu kategorii w języku informacyjno-wyszukiwawczym rzadko jest przedmiotem publikacji. Jeszcze rzadziej autorzy zajmują się związkiem pomiędzy obrazem świata zakodowanym w jiw, w tym za pomocą systemu kategorii, a obrazem utrwalonym w dokumentach i obrazem świata samych dokumentów. Najliczniejszą literaturę ma problematyka kategoryzacji rozumianej jako system kategorii pewnego języka informacyjno-wyszukiwawczego.

Mimo że celem pracy nie było stworzenie konkretnej kategoryzacji czy określenie natychmiastowych praktycznych implikacji kategoryzacji dla projektowania i używania języków informacyjno-wyszukiwawczych, to pewne elementy zaproponowanego podejścia zostały poddane weryfikacji praktycznej w czasie mojego kilkuletniego udziału w pracach nad tworzeniem centralnej kartoteki wzorcowej języka haseł przedmiotowych KABA w Centrum Formatów i Kartotek Haseł Wzorcowych Biblioteki Uniwersyteckiej w Warszawie. Wybrane tezy i fragmenty rozprawy prezentowałam na konferencjach (miedzy innymi: *Compatibility and integration of order systems. Research seminar of the TIP/ISKO meeting Warsaw, 13-15 September, 1995*; *Symposium nt. Opracowanie rzeczowe zbiorów bibliotecznych, Lublin 25-26 września 1998*; *Język haseł przedmiotowych KABA. Stan obecny i perspektywy rozwoju Sopot, 9-11 września 1997*), posiedzeniach Komisji ds. Opracowania Rzeczowego Zbiorów Stowarzyszenia Bibliotekarzy Polskich i seminariach naukowych (seminarium na temat teoretycznych podstaw języków informacyjno-wyszukiwawczych prowadzone przez profesor dr hab. Bożennę Bojar w Katedrze Lingwistyki Formalnej Uniwersytetu Warszawskiego; seminarium analizy systemów bibliotecznych prowadzone w Instytucie Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego przez profesor dr hab. Annę Sitarską). Pewne problemy poruszone w pracy przedstawiałam też studentom specjalizacji informacyjnej i studiów podyplomowych w moim macierzystym Instytucie Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego, słusznie licząc na ich dociekliwość.

Rozprawa mieści się w obszarze wiedzy szerszym niż teoria i praktyka języków informacyjno-wyszukiwawczych. Jej relewantność i użyteczność obejmuje teorię i praktykę systemów informacyjno-wyszukiwawczych, gdyż jiw i, co za tym idzie, zastosowana w nim kategoryzacja, nie istnieją w izolacji. Inaczej mówiąc, choć praca koncentruje się na właściwościach języków informacyjno-wyszukiwawczych, to w anali-

zach zostały uwzględnione różnorodne czynniki pozajęzykowe, tworzące substytutywne komponenty w przestrzeni komunikacji człowieka z systemem informacyjno-wyszukiwawczym. Systemowe, a nie typologiczne, widzenie jiw i rozważanej problematyki było intencją pracy, gdyż nadszedł czas eksponowania tego, co łączy, a nie tego, co dzieli różne obszary wiedzy w nauce o informacji i w innych naukach z nią powiązanych. Język informacyjno-wyszukiwawczy nie istnieje sam dla siebie. Jego celem nadrzędnym jako integralnego elementu systemu informacyjno-wyszukiwawczego<sup>15</sup> jest podnoszenie efektywności wyszukiwawczej siw. W tym kontekście jednym z ważniejszych czynników wydaje się być kategorialna organizacja systemu semantyczno-leksykalnego jiw, mająca z jednej strony wpływ na efektywność wyszukiwawczą języka, z drugiej – możliwość identyfikowania w języku relewantnych znaczeń. Języki prezentujące słownictwo w szeregu abecedowym, nawet jeśli wykorzystują struktury relacyjne, prowadzące do powstania mniej lub bardziej rozbudowanych drzew hierarchicznych, także języki będące systematyczną prezentacją pewnych podziałów dziedzinowych (dyscyplinarnych), w różny sposób ograniczają możliwości informacyjno-wyszukiwawcze systemu. Podczas gdy dobrze zbudowana kategoryzacja, oparta na rzetelnie określonych kryteriach relewancji – a nie jakakolwiek kategoryzacja – może te możliwości zwiększyć.

#### Konwencje typograficzne

- Wyrażenia języków naturalnych w supozycji materialnej pisane są kursywą, np. *informacja*;
- Ich znaczenia – czcionką tekstową i oznaczane „pojedynczym” cudzysłowem, np. ‘informacja’;
- Nazwy kategorii w językach naturalnych – dużymi literami, np. INFORMACJA.
- Cechy fonetyczne, semantyczne i atrybuty dystynktywne znaczenia – dużymi literami ujętymi w nawiasy kwadratowe, np. [ZENSKI].
- Wyrażenia języków informacyjno-wyszukiwawczych, nazwy kategorii, faset itp. zapisywane są z zachowaniem oryginalnej typografii. Wprowadzanie dodatkowych wyróżnień graficznych uznaję za uzasadnione jedynie w przypadkach, gdy ich brak mógłby oddziaływać na zrozumiałość tekstu. Wówczas wyrażenia jiw oznaczane są czcionką półgrubą.
- Nazwy kategorii w jiw, a dokładniej ważniejsze ich wystąpienia, wyróżnione są odmiennym krojem pisma.
- Wszystkie przyjęte konwencje typograficzne nie dotyczą tekstów cytowanych, w których zachowano oryginalne wyróżnienia graficzne.

---

<sup>15</sup> Ignorowanie kontekstu systemu w procesach projektowania i użytkowania języka prowadzi do fałszywego stereotypu dobrego (lub złego) języka informacyjno-wyszukiwawczego Por. Woźniak 1998.





## Podziękowania

*Serdecznie dziękuję wszystkim osobom, które słowem, myślą i czynem pomagały mi w mojej pracy.*

*Szczególny dług wdzięczności mam wobec Pani profesor dr hab. Bożenney Bojar, która nie tylko podjęła się trudu recenzowania pracy, ale zawsze wspierała mnie swoimi radami, wnikliwymi uwagami i ogromną życzliwością, wykazując mnóstwo zrozumienia dla moich słabości.*

*Dziękuję Panu profesorowi dr. hab. Marcinowi Drzewieckiemu, dyrektorowi Instytutu Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego, recenzentowi pracy, który zawsze miał dla mnie czas, udzielił wielu inspirujących rad i stworzył w Instytucie warunki sprzyjające pracy naukowej.*

*Dziękuję Pani profesor dr hab. Annie Sitarskiej i Panu profesorowi dr. hab. Janowi Wawrzyńczykowi, którzy zechcieli zapoznać się z pracą i podzielić ze mną swymi uwagami, które w istotny sposób przyczyniły się do ukształtowania ostatecznej jej postaci.*

*Dziękuję również wszystkim nie wymienionym tu z imienia osobom, które i tak wiedzą, że do nich się zwracam. Jestem Im wdzięczna i zobowiązana. Bez Ich cierpliwości, wyrozumiałości, rad i pomocy byłoby mi niewysłowienie trudno.*

Jadwiga Woźniak



## JĘZYK A REPREZENTACJA DOKUMENTU W SYSTEMIE INFORMACYJNO-WYSZUKIWAWCZYM

Proces tworzenia reprezentacji dokumentów w systemie informacyjno-wyszukiwawczym (siw) jest procesem używania pewnego języka w określonej sytuacji komunikacyjnej. Dlatego podstawą wszelkich teorii i działań praktycznych w tej dziedzinie powinna być teoria języka oraz pozostające z nią w związkach koncepcje i teorie psychologiczne, logiczne, antropologiczne, kognitywne i inne.

W Polsce refleksja teoretyczna nad reprezentacją dokumentów w systemach informacyjno-wyszukiwawczych i nad problemami języków informacyjno-wyszukiwawczych (jiw) osiągnęła wysoki poziom. Nadal istnieją jednak obszary, nawet w odniesieniu tylko do jiw stosowanych w systemach dokumentacyjnych, które wymagają dalszej pracy uwzględniającej bieżący dorobek nauki i interdyscyplinarny charakter wielu występujących tu problemów.

Rozdział pierwszy jest próbą sformułowania pewnych pytań dotyczących reprezentacji dokumentów w systemach informacji dokumentacyjnej oraz udzielenia odpowiedzi na niektóre z nich. Interesują nas tutaj, jak i w całej rozprawie, zagadnienia wyrażania informacji za pomocą środków języka informacyjno-wyszukiwawczego oraz sam język informacyjno-wyszukiwawczy. W zasadzie wyłączyliśmy z zakresu pracy problematykę dwóch pierwszych faz procesu indeksowania, tj. identyfikowania i selekcjonowania informacji, która po wyrażeniu jej w jiw utworzy reprezentację dokumentu lub pytania użytkownika, gdyż rzetelne i w miarę całościowe ujęcie tych zagadnień wymaga odrębnej rozprawy<sup>1</sup>.

Przez dziesięciolecia transformacja tekstu oryginalnego do postaci jego reprezentacji była przedmiotem zainteresowania nauki o informacji. Generalnie można wyróżnić dwa podstawowe podejścia do indeksowania i projektowania języków informacyjno-wyszukiwawczych – charakterystyki wyszuki-

---

<sup>1</sup> O problemach tych można przeczytać między innymi w publikacjach: Beghtol 1986; Bookstein 1975; Hjørland 1992; Soergel 1985; Wilson 1968.

wawcze dokumentów są albo derywowane z naturalnego języka indeksatora czy autora, np. poprzez automatyczne indeksowanie, albo są tworzone z wyrażeń pobranych z kontrolowanego słownika. Nadal niemal we wszystkich siw obowiązuje zasada tzw. najlepszego dopasowania (ang. *best match*), zakładająca że użytkownicy są zdolni do precyzyjnego charakteryzowania szukanej informacji, że zwerbalizowane potrzeby informacyjne są funkcjonalnymi ekwiwalentami tekstów dokumentów (informacji), że pytania informacyjne (zapytania, kwerendy) są odbiciem sytuacji problemowej użytkownika, że użytkownicy potrafią opisywać to, czego nie znają, że użyte przez nich wyrażenia wyszukiwawcze są właściwe itd. Praktyka nie potwierdza jednak prawdziwości tych założeń. Przy takim podejściu zainteresowanie skupia się głównie na dokumencie (i często na środkach technicznych systemu), w mniejszym stopniu na tym, co naprawdę się dzieje, gdy użytkownik i pracownik informacji prowadzą wyszukiwanie.

Charakterystyki wyszukiwawcze<sup>2</sup> w systemach dokumentacyjnych powstają w rezultacie transformacji treści<sup>3</sup> dokumentu i transformacji informacji pozat treściowych, co do których zakłada się, że mogą być relewantne dla potencjalnego użytkownika. David Blair<sup>4</sup> nazwał te dwie kategorie informacji odpowiednio informacjami treściowymi (tematycznymi, przedmiotowymi, rzeczowymi; ang. *subject*) i kontekstowymi (ang. *context*).

Mówiąc o informacji kontekstowej, zwykle wymienia się takie jej kategorie jak: autor, data publikacji, objętość dokumentu, jego język, forma wydawnicza

---

<sup>2</sup> Tworzenie charakterystyk wyszukiwawczych nie jest, co próbujemy uzasadnić w całej rozprawie, kwestią zbudowania jakiejś reprezentacji dokumentu, lecz takiej, która jak najlepiej będzie służyć wyszukiwaniu, na wybrani! spośród wielu możliwych sposobów reprezentacji tego, który w określonych warunkach jest najlepszy.

<sup>3</sup> Terminy *treść i przedmiot* (także *temat*) nic mają w informacji naukowej jednoznacznych sensów. „Pojęcie treści i przedmiotu książki zwyczajnie mieszało się, co utrudniało wielce uchwycenie istoty każdego z dwu rodzajów katalogu rzeczowego. W bibliologii dystynkcja ta dlatego jest ważna, że treść i przedmiot dzieł bynajmniej nie idą zawsze równolegle do siebie. Są książki wychodzące z jedncj zasady metodologicznej, mające zatem podobną treść, a traktujące o różnych przedmiotach, np. dzieło treści psychologicznej <<o zbrodni i karze>>, <<o powszechnych wyborach>> itp. Odwrotnie, o jednym przedmiocie traktują dzieła różnej treści, zależnie od tego, z jakiego stanowiska go rozpatrują, ku jakim cechom przedmiotu zwracają swe badania, do jakiego działu piśmiennictwa czy nauki należą sądy, wydane o tym przedmiocie. Np. o <<powszechnych wyborach>> mówi prócz psychologii – historia i socjologia i publicystyka; <<zbrodnia i kara>> bywa przedmiotem nauk prawnych, historii i socjologii, biologii i rozpatrywań religijnych, a nawet – tematem literatury pięknej”. Łysakowski 1928. Dokonując pewnego uproszczenia, przyjmijmy że przedmiotem dokumentu jest (omówiony w dokumencie) obiekt odpowiadający jakiemuś fragmentowi rzeczywistości; treścią zaś przynależne danemu przedmiotowi w danym dokumencie jego ujęcia, przedstawienia, interpretacje, lokalizacje itp., czyli jak pisał Ungurian, „zespół cech charakteryzujących tekst w sposób umożliwiający zakwalifikowanie dokumentu do określonej dyscypliny naukowej. Na cechy te składają się sądy i przedstawienia powzięte przez autora o danym przedmiocie (co i jak autor myśli), jego psychiczne i społeczne nastawienie, intencje: naukowe bądź publicystyczne, stawiane sobie cele i zadania, nazewnictwo specyficzne dla określonej dyscypliny oraz styl”. Ungurian 1976, s. 77.

<sup>4</sup> Blair 1990.

itd. Niekiedy projektując język informacyjno-wyszukiwawczy, zapomina się, że różne dokumenty wymagają zróżnicowanej informacji kontekstowej. Na przykład w systemie informacyjno-wyszukiwawczym dla rozgłośni radiowej zapewne największą część zbioru wyszukiwawczego będą stanowiły reprezentacje typowych dla tego miejsca dokumentów, tj. nagrań audycji radiowych i muzyki. Kategorie informacji kontekstowej przeniesione z bibliotek nie sprawdzą się w tym środowisku wyszukiwawczym, bo inna jest specyfika charakteryzowanych tu dokumentów i inne potrzeby użytkowników tego systemu. Podobna sytuacja może mieć miejsce w bibliotece. Różne typy dokumentów gromadzonych przez biblioteki i różne oczekiwania użytkowników wymagają różnych kategorii kontekstu. Wiesław Babik<sup>5</sup> na podstawie analizy materiałów kartograficznych wyróżnił następujące kategorie informacji<sup>6</sup> kontekstowej: rodzaj materiału kartograficznego, kształt, nośnik obrazu (przedstawienia) kartograficznego, format, forma wydawnicza, przeznaczenie, język opisu, autorstwo, wydawca, skala, odwzorowanie, siatka kartograficzna, układ współrzędnych, orientacja. W kategorii informacji kontekstowej mieszczą się również informacje na temat środowiska wyszukiwawczego, w którym dokument istnieje, na przykład mówiące o liczbie wyszukiwań danego dokumentu, jego wypożyczeń, datach wprowadzenia informacji o dokumencie do bazy itd.

Informacje kontekstowe są relatywnie proste do wyrażenia w charakterystyce dokumentu, gdyż cechy kontekstowe charakteryzują się binarnością, tzn. dokument albo posiada daną cechę, albo jej nie posiada. Przyczyn pojawiających się niekiedy problemów z określeniem wartości cechy w odniesieniu do konkretnego dokumentu należałoby upatrywać raczej w nieostrości lub niewyraźności<sup>7</sup> wyrażenia cechy, a nie w samych dokumentach. Trudniejsze jest przewidzenie, jaki wpływ na efektywność wyszukiwania informacji będą miały zaprojektowane lub błędnie pominięte kategorie informacji kontekstowej. Tradycyjnie jednym z ważnych, choć rzadko i w ograniczonym zakresie wykorzystywanych w praktyce, źródeł informacji kontekstowej są badania użytkowników i zapytań informacyjnych, zwłaszcza ich wersji finalnych będących podstawą wyszukiwania. Pewną poprawę sytuacji przyniosły badania kognitywne w nauce o informacji, a wśród nich badania organizacji wiedzy użytkowników. Ponieważ problematyka kognitywna nie była jak dotąd przed-

<sup>5</sup> Babik 1992.

<sup>6</sup> Kategorie informacji rzeczowej tworzą tu: dziedziny wiedzy, gałęzi gospodarki narodowej, sfery działalności ludzkiej, obiekty, zjawiska, procesy, atrybuty obiektów, atrybuty zjawisk, atrybuty procesów. Babik 1992, s. 59.

<sup>7</sup> „Nicostrość wyrażenia powoduje kłopoty z jego zakresem, a ściślej kłopoty przy podejmowaniu decyzji czy dany przedmiot zaliczyć do tego zakresu. Tymczasem niewyraźność jest <<mankamentem>> znaczenia, polegającym na tym, że tylko o pewnych cechach wiadomo, iż do niego należą, i tylko o pewnych wiadomo, że nie należą; ale są i takie, co do których nie sposób rozstrzygnąć, czy stanowią elementy tego znaczenia, czy też nie stanowią. Nie sposób zaś rozstrzygnąć nic na skutek nieznajomości języka, tego, iż użytkownik języka niewyraźnie rozumie dane wyrażenie, choć ma ono sprecyzowane znaczenie, lecz na skutek tego, iż owo znaczenie jest nie sprecyzowane”. Pełc 1984, s. 183-184.

miotem publikacji w języku polskim, dlatego zdecydowaliśmy się poświęcić jej fragment rozdziału, aby wzbogacić tło, na którym chcemy pokazać i uzasadnić potrzebę modyfikacji podejścia do projektowania języków informacyjno-wyszukiwawczych, zwłaszcza ich aspektów kategoryzacyjnych.

W systemach dokumentacyjnych budowanie charakterystyk wyszukiwawczych w oparciu o informacje treściowe polega na przydzieleniu dokumentowi pewnej liczby wyrażeń mówiących „o czym” i ewentualnie „dla kogo” jest dokument. Nie jest to tak proste, jak określanie przynależności do poszczególnych kategorii kontekstowych. Pokazują to między innymi badania zgodności indeksowania<sup>8</sup>, z których wynika, że prawdopodobieństwo wskazania przez różnych indeksujących tego samego wyrażenia jako relewantnego dla danego dokumentu rzadko bywa większe niż 75%, a na ogół jest niższe<sup>9</sup>. Nie wiadomo też, jak zgodność indeksowania, a właściwie jej brak, wpływa na efektywność wyszukiwania, gdyż w zasadzie nie ma publikacji na ten temat. Pewną ilustracją wagi omawianego problemu jest nadal trwająca w literaturze dyskusja nad pojęciem i naturą przedmiotu dokumentu. Za Hjørlandem<sup>10</sup> omówimy to zagadnienie, wyróżniając cztery podejścia do pojęcia przedmiotu: idealizm subiektywny, idealizm obiektywny, pragmatyczne ujęcie przedmiotu i realistyczna (materialistyczna) teoria przedmiotu.

## PRZEDMIOT DOKUMENTU

Wszelkie idealistyczne ujęcie przedmiotu oznacza, że jest on desygnatem pewnej idei. Pojęcia reprezentują coś „uniwersalnego, co istnieje poza ludzką świadomością i niezależnie od niej, a co jednocześnie jest aprioryczne w stosunku do poszczególnych rzeczy (pierwotnie w powiązaniu z Bogiem, obecnie raczej jako forma poznania a priori w sensie Kanta). W przełożeniu na kategorię przedmiotu oznacza to, że konkretne dokumenty <<partycypują>> w ideach wyrażonych w danym przedmiocie. Idee te istnieją poza ludzką świadomością (lub w jej ramach jako postrzeżenia a priori), a ponadto są pierwotne w stosunku do indywidualnych pojęć wyrażonych w pojedynczych dokumentach. Idee te czy przedmioty posiadają cechy uniwersalne lub stałe; mogą być analizowane jednorazowo i zbiorczo w ramach uniwersalnego systemu bądź też wydzielone w pojedynczych częściach”<sup>11</sup>.

Idealizm subiektywny przyjmuje, że pojęcia są wyrazem percepcji lub poglądów jednego czy wielu podmiotów, są tym, co podmioty te subiektywnie pojmują czy rozumieją. Klucz do pojęcia przedmiotu znajduje się zatem w badaniu umysłów. W przypadku dokumentów bibliotecznych istnieje co najmniej kilka możliwych interpretacji tak rozumianego przedmiotu: wersja

---

<sup>8</sup> Cleverdon 1970, 1967, 1964.

<sup>9</sup> Zunde 1969.

<sup>10</sup> Hjørland 1992.

<sup>11</sup> Ibidem.

autora (często w postaci wyrażonej *explicite* lub *implicite* w tytule lub tekście, zwłaszcza w streszczeniu autorskim), wersja użytkownika (zwykle tym bardziej zróżnicowana, im więcej jest użytkowników), wersja wydawcy (wskazana na przykład w nazwie serii wydawniczej), wersja bibliotekarza, która może być wyrażona w kategoriach języka informacyjno-wyszukiwawczego.

Niektórzy teoretycy wyszukiwania informacji wydają się wychodzić z założenia, że system wyszukiwania informacji powinien systematyzować przedmioty według subiektywnego ujęcia każdego użytkownika<sup>12</sup>. Wykazują przy tym skłonność do opierania się na psychologicznych badaniach postrzegania przedmiotów przez użytkowników, na ich strukturach wiedzy. Trudno nie zgodzić się ze stwierdzeniem, że istnieją pewne rodzaje systemów informacyjnych, które powinny zmierzać do dostosowania opisu przedmiotów do subiektywnych potrzeb użytkownika – na przykład systemy biblioteczne dla dzieci lub systemy dydaktyczne, w których punkt wyjścia i cel można opisać zarówno ze względu na potrzeby procesu uczenia się, jak i wspierania radą uczniów. Oba typy systemów wyrażają pewien rodzaj paternalizmu, polegającego na przyjęciu przez kogoś odpowiedzialności za poszukiwanie informacji przez innych. Dzieje się to poprzez założenie tworzenia powiązań między dokumentami i uniwersum przedmiotowym użytkownika, a więc poprzez podjęcie się interpretacji przedmiotów, czy też zawartości informacyjnej dokumentów z punktu widzenia psychologicznej bądź pedagogicznej oceny potrzeb i celów. „Czy jednak, niezależnie od takiego paternalistycznego podejścia, powinno się uwzględniać psychologię użytkownika przy opisie przedmiotowym? Tak, na swój sposób jest to rzeczywiście pożądane. Systemy informacyjno-wyszukiwawcze powinny być przyjazne dla użytkownika, a można to osiągnąć dzięki znajomości języka użytkownika i jego subiektywnej percepcji oraz poprzez wykorzystanie tej wiedzy np. w systemie odsyłaczy do terminów zalecanych. Zadanie powiązania wszystkich systemów z użytkownikami oznaczałoby prawdopodobnie pewien ideał. Nie oznacza to jednak interpretacji treści dokumentów na podstawie subiektywnych odczuć użytkowników, lecz wykorzystanie percepcji użytkownika do stworzenia odsyłaczy i instrukcji, czyli do nadania systemowi cechy przyjazności. Moim zdaniem, zagadnienie przyjazności systemu nie jest centralnym problemem teoretycznym w wyszukiwaniu informacji. Głównym problemem jest reprezentacja wiedzy zawartej w dokumentach. Problem przyjazności systemu jest kwestią kognitywno-ergonomiczną, która musi być rozwiązana w systemie, lecz ma ona znaczenie drugoplanowe w porównaniu z jakością reprezentacji wiedzy w bazach danych. Systemy informacji naukowej muszą – w moim przekonaniu – opierać się na założeniu, że to użytkownik przyswaja kategorie, terminologię i klasyfikacje nauki, wiedzę oraz poznaje systemy informacji, a nie odwrotnie. Przyswojenie

---

<sup>12</sup> Na przykład: Salton 1970; Belkin 1982.



przez naukę i jej systemy informacji kategorii i terminologii użytkownika jest zadaniem działalności popularyzacyjnej, a nie podstawowym zadaniem informacji naukowej. Często można spotkać się z postulatem stosowania zasad psychologii i psycholingwistyki w projektowaniu systemów, lecz zapomina się, że zasady te stwarzają dylematy czy wręcz sprzeczności (...). Wniosek wynikający w tym miejscu jest następujący: ten kto poszukuje klucza do pojęcia przedmiotu w umyśle użytkownika popełnia błąd psychologizmu”<sup>13</sup>.

Typowe subiektywno-idealistyczne ujęcie przedmiotu polega zatem na eksponowaniu pewnych jego własności bądź z punktu widzenia autora, bądź czytelnika, bądź innego interpretującego. Jak dotąd nie stwierdzono, by subiektywne podejście do dokumentu mogło zapewnić optymalną analizę jego treści. Być może stało się to przyczyną agnostycznej koncepcji przedmiotu – nie można stwierdzić, co jest przedmiotem i jak należy go określić. Pogląd taki sformułował między innymi Patrick Wilson w *Two kinds of power*. „Patrick Wilson bada – zwłaszcza poprzez eksperymenty myślowe – przydatność różnorodnych metod ustalania przedmiotu dokumentu. Na metody te składa się: 1. Ustalanie celu przyjętego przez autora przy pisaniu dokumentu; 2. Oszacowanie dominacji lub podporządkowania różnych elementów w obrazie powstałym w wyniku czytania dokumentu; 3. Grupowanie lub wyliczenie wykorzystania w dokumencie pojęć i odesłań; 4. Ustalenie zbioru reguł wyboru tzw. zasadniczych elementów w całości dokumentu (w przeciwieństwie do elementów nieznaczących). Przekonująco wykazuje, że żadna z tych metod nie jest wystarczająca do ustalenia przedmiotu dokumentu, kończąc konkluzją: pojęcie przedmiotu tekstu jest nieokreślone (...). W związku z tą ostatnią uwagą Wilson włącza interesujący przypis, w którym zwraca uwagę na częste nieprecyzyjne stosowanie pojęć przez autorów dokumentów (jako przykład podaje „hostility” [„wrogość” – przypis JW]). Nawet jeśli bibliotekarz może uchwycić bardzo dokładne znaczenie danego pojęcia, to nie będzie w stanie go wykorzystać w swojej klasyfikacji, ponieważ żaden z dokumentów nie stosuje tego pojęcia równie precyzyjnie. Dlatego też, konkluduje Wilson: jeśli ludzie piszą o zjawiskach dla nich mętnie zdefiniowanych, to opis tych przedmiotów musi tę mętność odzwierciedlać”<sup>14</sup>.

Agnostycyzm, jak ten wyrażany przez Wilsona, nie zniechęcił innych autorów do szukania odpowiedzi na pytanie o przedmiot dokumentu. Wielu z nich poszło jednak inną drogą, zakładając, że nie jest możliwe ustalenie przedmiotu poprzez badanie światów mentalnych autorów, użytkowników czy jakichkolwiek innych grup osób. Autorzy ci szukali istoty przedmiotu w innych sferach bytów. Naturalną kosekwencją takiej postawy było pytanie o obiektywne kryteria ustalenia przedmiotu dokumentu.

---

<sup>13</sup> Hjørland 1992, s. 175.

<sup>14</sup> Ibidem s. 176.

Subiektywno-idealistyczna teoria przedmiotu traktuje go jako kategorię subiektywną. Za nieistotne przyjmuje się, że te subiektywne kategorie mogą być mniej lub bardziej powtarzalne i identyczne u różnych osób; zasada wyraża się w tym, że są one zależne od subiektywnych wyobrażeń. Idealizm obiektywny nie traktuje przedmiotu obiektywnie na zasadzie, że różne osoby dojdą do tego samego przedmiotu danego dokumentu, jeśli przeprowadzą poprawną analizę, a przedmiot dokumentu następnie będzie można uznać za obiektywny (przynajmniej w szczególnym znaczeniu tego słowa). Jeśli idealizm subiektywny cechuje się nadmiernym eksponowaniem percepcji zmysłowej, to idealizm obiektywny skłonny jest do nadmiernego eksponowania pewnych aspektów analizy teoretycznej i uznawania ich za absolutne. Pogląd, według którego, myśl i świadomość, język naturalny oraz uniwersum przedmiotowe ludzi są zasadniczo niezależne od funkcjonalnego kontekstu procesów mentalnych, stanowi wzorzec idealistycznego ujęcia przedmiotu. Wzorzec ten pozostaje w prostej opozycji do poglądu, że procesy mentalne są narzędziem ukształtowanym przez zadania i warunki, w których zachodzą i do nich dostosowanym. Dalszą konsekwencją idealistyczno-obiektywnego ujęcia przedmiotu jest nieuznawanie ani światopoglądowych, ani dziedzinowych czy jakichkolwiek innych preferencji, co pozostaje w ewidentnej sprzeczności nawet z pobieżnym oglądem rzeczywistości bibliotek i systemów informacyjno-wyszukiwawczych.

Ani postawa idealistyczno-subiektywna, ani idealistyczno-obiektywna nie uwzględniały między innymi aspektu potencjalnego wykorzystania dokumentów. Lukę tę stara się zapełnić kolejna koncepcja – pragmatyczne ujęcie przedmiotu dokumentu. Założeniem tej koncepcji jest przekonanie, że zapotrzebowanie na informację jest pochodną problemu, wobec rozwiązania którego staje użytkownik. Aby zapis przedmiotów sporządzony przez bibliotekarzy czy pracowników informacji był sensowny, musi antycypować potrzeby użytkowników – musi umożliwiać użytkownikowi znalezienie tego, czego on poszukuje. Informacja przedmiotowa pełni zatem funkcję instrumentalną lub pragmatyczną – dokumenty są indeksowane w celu ich wyszukania w odpowiedzi na określoną sytuację problemową użytkownika.

Dagobert Soergel<sup>15</sup> wprowadził rozróżnienie między indeksowaniem zorientowanym na treść (ang. *content oriented*) i indeksowaniem zorientowanym na zapytanie (ang. *request oriented*). Hjørland podaje przykład bazy danych *Ringdok*, w której literatura chemiczna opisywana jest inaczej niż w *Chemical Abstracts*, ponieważ *Ringdok* jest ukierunkowana na obsługę przede wszystkim potrzeb przemysłu farmaceutycznego. „Indeksowanie ukierunkowane na treść jest opisem przedmiotów, który ma być traktowany wyłącznie jako funkcja atrybutów danego dokumentu: tak, jak w stwierdzeniu <<dokument ten zawiera wzór chemiczny kwasu siarkowego>> (co w konsekwencji prowadzi do

---

<sup>15</sup> Cyt. za Hjørland 1992, s. 178.

klasyfikacji «chemia nieorganiczna»). Indeksowanie ukierunkowane na użytkownika czy też na zapytanie informacyjne jest opisem przedmiotu, który należy postrzegać jako relację między cechami dokumentu oraz rzeczywistymi lub antycypowanymi potrzebami użytkownika. «Niniejszy dokument poświęcony jest kwasowi siarkowemu. Kwas siarkowy ma działanie żrące. Grawerzy potrzebują odczynników żrących» – i stąd powstaje przyporządkowanie, np. «literatura na temat użycia chemikaliów w grawerstwie». Indeksowanie ukierunkowane na potrzeby jest relacją instrumentalną (środek – cel) między dokumentem i potrzebą użytkownika<sup>16</sup>. Nie jest możliwe, aby w bibliotece uniwersalnej scharakteryzowano choćby przytoczony przykład dokumentu o kwasie siarkowym pod względem wszystkich potencjalnych zastosowań. A zatem koncepcja Soergela dotycząca indeksowania ukierunkowanego na zapytania, choć istotna i twórcza, jest w praktyce możliwa do wykorzystania jedynie w ograniczonym zakresie, w specjalistycznych systemach informacyjnych.

Pragmatyczna teoria przedmiotu, mając swoje ograniczenia, zdaje się stanowić znaczący wkład w postrzeganie przedmiotu dzięki wyeksponowaniu jego zadaniowego, celowościowego i instrumentalnego charakteru, przedmioty przestają być immanentnymi jakościami, stają się raczej wartościami rzeczy.

Ostatnim z podejść jest realistyczna (materialistyczna) teoria przedmiotu. Według światopoglądu realistycznego i materialistycznego obiektywnie istniejące rzeczy mają obiektywne własności. „Dokument, posiadając pewne obiektywne własności, odzwierciedla jednocześnie subiektywne ujęcie przedmiotu przez autora. Jeśli dokument stwierdza, że «inteligencja jakiejś osoby jest skorelowana z wielkością jej mózgu», to jest to sąd subiektywny (i fałszywy). Jednakże obiektywnym faktem jest to, że dokument zawiera ten (fałszywy) sąd. Interesują nas tutaj obiektywne własności dokumentów. Te obiektywne własności nie są ani subiektywnymi sędziami, ani też ocenami zawartymi w dokumentach; obiektywne własności posiadają potencjał kognitywny (lub informacyjny) – zakładamy, że czytelnik rozróżnia twierdzenia prawdziwe i fałszywe<sup>17</sup>”.

Teorię materialistyczną cechuje pewien związek z epistemologią. Realistyczna i materialistyczna teoria przedmiotu nie ogranicza się jedynie do rozwiązywania doraźnie traktowanych problemów, lecz dąży do uświadomienia długofalowych konsekwencji przyjętego podejścia. U jej podstaw leży przekonanie, że przedmiotów nie należy systematyzować w ograniczony, instrumentalny sposób, lecz podejmować próby przyczyniania się do wnikliwego penetrowania przez naukę istoty rzeczy. Kategorie semantyczne powinny odzwierciedlać istotne i ogólne cechy rzeczywistości. W praktyce będą to często pojęcia naukowe. Naturalnie, nauka nie jest ani niekontrowersyjna, ani obiektywna, ani nieomylna, lecz przynajmniej dyskusja o obiektywności badań nauko-

---

<sup>16</sup> Hjørland 1992, s. 180.

<sup>17</sup> Ibidem s. 182-183.

wych stanowi jej część. Stąd też sama analiza przedmiotu, zależna od czynników kontekstowych, do których należy wielkość istniejącej literatury oraz system dostępu do niej, jest w istocie częścią procesu zdobywania wiedzy.

Według Hjørlanda obiektywność w odniesieniu do dokumentu może oznaczać: (1) niezależność podmiotu, który przyswaja tekst; (2) zgodność z rzeczywistością. W pierwszym znaczeniu im więcej czytelników przypisuje dokumentowi te same cechy, tym większy jest obiektywizm. W drugim przypadku (zgodność z rzeczywistością) relacja ta jest odwrotnie proporcjonalna. Ponieważ do ustalenia znaczących cech tekstu naukowego niezbędne są specjalistyczne kwalifikacje, przypuszczalnie niewielka grupa osób jest w stanie uchwycić pełny potencjał (treści) pracy. Innymi słowy, cechy bez trudu ustalone przez wiele osób, często są mniej znaczące (lub bardziej przypadkowe), a przez to mniej obiektywne w drugim znaczeniu. Przy ocenie najbardziej znaczących cech książki, a zatem i jej przedmiotów, istnieje opozycja między tymi dwoma ujęciami. Twierdzenie, że cechy dokumentu mają obiektywny charakter, chociaż ich opis wymaga pewnych subiektywnych warunków wstępnych, implikuje, że to rzeczywistość, weryfikacja dokumentu w praktyce, ostatecznie decydują o jego informacyjnym potencjale, niezależnie od tego, ile chybionych ujęć wcześniej sformułowano. Różne cechy dokumentów mogą mieć różne znaczenie dla rozmaitych celów czy dyscyplin naukowych. Dyscypliny czy teorie naukowe mogą być różnie umiejscowione epistemologicznie. Stąd mogą wynikać różnice we wskazywanych głównych cechach dokumentów. Identyfikacja cech z wąskiego teoretycznego punktu widzenia jest bardziej pragmatyczna niż dokonywana z bardziej ogólnej perspektywy. Identyfikacja cech dokumentów z nadrzędnego czy ogólnego punktu widzenia presuponuje umiejętność oceny potencjalnych elementów różnych teorii.

Ujmowanie przedmiotu jako funkcji cech dokumentu nie mówi jeszcze, czym jest przedmiot. Aby określić pojęcie przedmiotu należałoby skoncentrować się na ustaleniu, które cechy dokumentu należą do opisu przedmiotu i w jaki sposób spełniają swoją rolę. Pozornie stwierdzenie, co jest przedmiotem dokumentu, jest proste – wymaga wskazania kilku znaczących cech dokumentu, a raczej rzeczywistości, którą dokument odzwierciedla. Jak pisze Hjørland, jeśli dokument traktuje o stylu budownictwa za Christiana IV, to można mu przypisać jako przedmiot „styl budownictwa za Christiana IV”. Wciąż nie jest jednak jasne, dlaczego w danym przypadku właśnie ta a nie inna cecha została wybrana jako przedmiot. „Przedmioty same w sobie należy zdefiniować jako jednostki epistemologicznego potencjału dokumentu. Potencjał jest raczej własnością nieuchwytną – taki jest też problem z określaniem przedmiotów. Jednakże potencjał rzeczy nie jest subiektywną czy obiektywną <<idea>>. Potencjał jest obiektywną możliwością. Uran posiadał swój potencjał jako paliwo jądrowe zanim nauka uświadomiła te możliwości, a wielu autorów zmarło zanim uznano znaczący potencjał ich pracy. O potencjale rzeczy i prac decyduje aktualny etap rozwoju społeczeństwa. Na jednym

etapie uran bez swojego potencjału nie jest uważany za szczególnie wartościowy metal. Na drugim etapie jest ważnym źródłem energii, a na kolejnym być może okaże się jeszcze czymś innym. Trzeba więc stwierdzić, że to poziom rozwoju społeczeństwa i praktyka są tym, co tworzy przedmiot”<sup>18</sup>.

Zatem opis przedmiotowy dokumentu jest wykładnikiem epistemologicznego potencjału dokumentu, tak jak widzi go ten, kto opisuje przedmiot. Im dokładniejszy jest przewidywany opis potencjału dokumentu, tym wartościowszy jest opis przedmiotu. Postrzeganie potencjału dokumentów zależy od uprzedniego rozumienia przedmiotu przez osobę, która zajmuje się jego określeniem. Hjørland jest zwolennikiem pojęcia obiektywnego potencjału czy też obiektywnych przedmiotów dokumentów. Opis przedmiotu jest zatem prognozą przyszłego potencjału. Prognoza ta może opierać się zarówno na sądach pozytywnych, jak i negatywnych. Na opis przedmiotu można patrzeć równocześnie jako na rodzaj wizji oraz ocenę w odniesieniu do bieżących badań. Najważniejszym warunkiem w opisie przedmiotu nie jest szczególna metoda, lecz dojrzałość osądu indeksatora.

Opis przedmiotu rzadko prezentowany jest jako bezpośredni opis potencjału dokumentu; znacznie częściej ma postać odesłania do jakiejś dyscypliny naukowej (np. „przedmiotem jest psychologia”), a więc zdefiniowanego społecznie obszaru problemowego, w ramach którego dokument wnosi szczególny wkład w rozwiązywanie problemów. Przedmioty dokumentów można wyrażać także nie wprost, jedynie sygnalizując ich szczególne właściwości (np. „poświęcony jest architekturze za Christiana IV”), które w dalszej kolejności mogą odsyłać do dyscypliny naukowej lub też mogą bezpośrednio stanowić podstawę oceny przedmiotu dokumentu przez użytkownika (np. „atrakcje turystyczne”).

Zdaniem Hjørlanda celem analizy przedmiotu jest określenie, czy dokument reprezentuje potencjał epistemologiczny w odniesieniu do przyszłych użytkowników danej kategorii, czy w odniesieniu do danego pojęcia (por. przykład pojęcia *hostility*). Przypisanie dokumentowi przez indeksatora określonego przedmiotu oznacza wypowiedzenie sądu „ten dokument posiada potencjał epistemologiczny w zakresie...”. Określenie potencjału epistemologicznego zależy zatem nie tylko od stanu nauki czy szerzej wiedzy w określonym czasie, ale także od stanu różnych segmentów wiedzy indeksatora w tym wiedzy ogólnej, wiedzy o temacie, użytkowniku, jego problemie wyszukiwawczym i samym systemie informacyjno-wyszukiwawczym. Właściwie to wiedza indeksatora decyduje o tym, czy potencjał dokumentu zostanie prawidłowo zidentyfikowany i nazwany.

---

<sup>18</sup> Hjørland 1992, s. 185.

## TWORZENIE REPREZENTACJI DOKUMENTU – ASPEKTY JĘZYKOZNAWCZE

Tworzenie reprezentacji dokumentu jest rodzajem działania lingwistycznego ze wszystkimi jego możliwościami i ograniczeniami. Dorobek językoznawstwa pokazuje, jak niełatwa jest odpowiedź na pytanie o znaczenie<sup>19</sup> (znaczenie *znaczenia* i konkretnego słowa). Podobnie jest z pytaniem o znaczenie wyrażenia użytego do opisu treściowego oraz samych pojęć opisu i wyszukiwania treściowego. W rozprawie w zasadzie nie będziemy się zajmować istotą znaczenia. Większą wagę ma dla nas problem, jak efektywnie przedstawiać i wyszukiwać znaczenia w języku informacyjno-wyszukiwawczym. Przyjmujemy też, że czytelnicy rozumieją, co ma się na myśli, mówiąc w kontekście dokumentacyjnego systemu informacyjno-wyszukiwawczego o opisie czy wyszukiwaniu treściowym (rzeczowym).

Języki informacyjno-wyszukiwawcze są tworzone na bazie różnych referencjalnych koncepcji znaczenia<sup>20</sup>. Najczęściej projektanci jiw wykorzystują koncepcję znaczenia jako przedmiotu (klasy przedmiotów) lub cechy przedmiotu, czyli znaczenie denotacyjno-konotacyjne. Spośród różnych ujęć znaczenia obiecująca wydaje się, dotychczas niebrana pod uwagę w projektowaniu jiw, areferencjalna koncepcja znaczenia jako użycia<sup>21</sup>. „Ujęcie to, charakterystyczne dla tzw. oksfordzkiej filozofii języka ostatnich trzydziestu lat, realizuje postulat sformułowany przez jednego z twórców tej szkoły, L. Wittgensteina, który w *Dociekaniach filozoficznych* pisał: «Nie szukajcie znaczenia, szukajcie użycia (wyrażeń)» G. Ryle zaś dodawał: «Znać znaczenie wyrażenia to znać zasady jego użycia», tzn. wiedzieć, w jakich warunkach i okolicznościach można danego wyrażenia użyć. Opis użyć to opis znaczenia”<sup>22</sup>. „Wittgenstein nawoływał, by nie pytać o znaczenie wyrażeń, lecz zająć się ich użyciami, to jest przyglądać się, jak przebiega «gra językowa», złożona z językowych i niejęzykowych zachowań rozmówców. Na podstawie

---

<sup>19</sup> Trudno jest tu podać choćby wybór literatury ze względu na jej niezwykłą liczebność. Problem znaczenia był i jest przedmiotem namysłu i rozważań od ponad dwóch i pół tysiąca lat. Dlatego rezygnujemy choćby z próby wskazania ważniejszych publikacji, odsyłając czytelnika do Bibliografii załącznikowej, rejestrującej prace, wykorzystane przy pisaniu rozprawy.

<sup>20</sup> Referencjalne koncepcje znaczenia sytuują znaczenie na zewnątrz języka, uznając, że jest nim jakiś pozajęzykowy odpowiednik wyrażenia bądź też związek między wyrażeniem a zjawiskami zewnętrznymi.

<sup>21</sup> Rozumienie znaczenia jako sposobu użycia znaku językowego pochodzi od Wittgensteina i jego kontynuatorów oraz (niezależnie) od Kazimierza Ajdukiewicza. Podstawowym pojęciem u Ajdukiewicza jest pojęcie językowych dyrektyw uznawania zdań, a więc stosowania zdań w określonych okolicznościach. Teoria Ajdukiewicza ogranicza się do badania wyrażeń będących zdaniami w sensie logicznym (tylko wobec takich można zajmować postawę uznawania) lub składnikami takich zdań, z pominięciem wypowiedzi ekspresyjnych, rozkazów, przyrzeczeń itp., które w teoriach z kręgu Wittgensteina są również przedmiotem analizy. Inaczej mówiąc, znaczenie definiowanym przez Ajdukiewicza jest tylko znaczenie poznawcze.

<sup>22</sup> Grzegorzczkova 1995, s. 19.

przebiegu tej gry można odtworzyć reguły, które nią rządzą, reguły używania wyrażań. To właśnie owe reguły posługiwania się nazwą w taki sposób, by oznaczała pewne rzeczy, a zdaniem w taki sposób, by – prawdziwie albo fałszywie – zdawało sprawę z pewnego zdarzenia, pytało albo rozkazywało – to właśnie owe reguły gry językowej mamy na myśli mówiąc o znaczeniach wyrażań. Ale owe znaczenia nie są żadnymi rzeczami posiadanymi przez te wyrażenia. Mówimy wprawdzie, że np. zdanie «podaj cegłę» ma znaczenie, ale nie chodzi nam o nic innego jak o to tylko, co zdoła wywnioskować obserwujący – zwykłą w czasach Wittgensteina – scenę na budowie: majster murarski krzyczy «podaj cegłę», a pomocnik na ten rozkaz ją podaje. Owo zdanie, gesty oraz towarzysząca im sytuacja – oto tzw. znaczenie słów «podaj cegłę»<sup>23</sup>. „Wielką klasę przypadków stosowania słowa «znaczenie» – choć nie wszystkie jego zastosowania – można objaśnić tak: znaczeniem słowa jest sposób użycia go w języku”<sup>24</sup>. Nie jest naszym celem ani dowiedzenie poprawności powyższego sformułowania, ani jego obrona przed zwolennikami innych koncepcji. Wydaje się jednak wysoce prawdopodobne, że zarówno samo pojęcie reprezentacji dokumentu, jak i elementów tworzących reprezentację, tj. różnych poziomów znaki jiw, należą do tej „wielkiej klasy przypadków”, w których znaczenia mogą być objaśnione przez pokazanie użycia. Pozostawiając kwestię znaczenia znaków języka naturalnego językoznawcom i filozofom języka, rozważanie koncepcji znaczenia jako użycia ograniczymy do tego, co jest specyficzne tylko dla języków informacyjno-wyszukiwawczych, tj. do znaczenia jako użycia znaków jiw w odniesieniu do rzeczywistości dokumentacyjnej. Naszym zdaniem ujmowanie znaczenia jako użycia może i powinno mieć zastosowanie w projektowaniu jiw jako metoda komplementarna, dopełniająca, nie zastępująca, tradycyjnie stosowane podejście denotacyjne i konotacyjne. Dzięki takiemu podejściu opis znaczenia w jiw można byłoby wzbogacić o dodatkowy element, jakim jest użycie znaku jiw w charakterystyce konkretnego dokumentu w kontekście innych znaków tego języka.

Wittgenstein przyrównuje znaki języka do narzędzi, a znaczenia znaków do funkcji pełnionych przez narzędzia. Wszystkie narzędzia służą do wywoływania zmian w przedmiotach; znaki językowe są narzędziami służącymi do wywoływania zachowań. W koncepcji Wittgensteina znaczenie nie ma desygnatu w żadnej dziedzinie (ani psychicznej, ani obiektywno-realnej, ani obiektywno-idealnej), ponieważ wbrew swej gramatycznej formie nie jest nazwą; jest fragmentem takich zwrotów jak „x ma znaczenie” lub „x i y mają to samo znaczenie” (tzn. wywołują takie same zachowania). „Dla problemu znaczenia omawiany kierunek badawczy okazał się bardzo ważny. Przede wszystkim odwrócono się całkowicie od ujęcia denotacyjnego, tak bardzo zakorzonego w tradycji logicznej i lingwistycznej. (...) Ryle doprowadzał do absurdu przekonanie, że znaczeniem wyrażenia: *pierwszy człowiek, który zdobył Mount*

---

<sup>23</sup> Pełc 1984, s. 213-214.

<sup>24</sup> Wittgenstein 1972, s. 34.

*Everest* może być Hilary, człowiek oddychający, nagrodzony przez królową. Pisze on: «Znaczenia to nie rzeczy, ani nawet – bardzo dziwne rzeczy. Poznanie znaczenia wyrażenia jest bardziej podobne do uczenia się musztry aniżeli do natrafiania na dotychczas nie spotykany obiekt. Polega ono na opanowaniu sposobów poprawnego operowania danym wyrażeniem oraz wszelkimi jego równoważnikami»<sup>25</sup>. Implikacją takiej koncepcji jest między innymi włączenie do znaczenia elementów sytuacyjnych, wynikających z użycia znaku językowego. Jeżeli znaczenie wyprowadza się z użycia, wobec tego do znaczeń należą typowe składniki sytuacji, na przykład znaczenie zwrotu *Jak się masz?* zawiera „powitanie”, znaczenie wyrażenia *zły pies* na tabliczce na bramie zawiera „ostrzeżenie” itp. Do znaczenia użytych znaków języka wprowadza się także intencje nadawcy, często ukryte i niewidoczne na pierwszy rzut oka.

Dotychczasowa praktyka języków informacyjno-wyszukiwawczych charakteryzuje się tym, że opis znaczenia, zwykle interpretowanego denotacyjno-konotacyjnie, w jiw albo jest niejawny, tzn. użytkownik języka na podstawie wyrażonych w słowniku jiw relacji ekwiwalencji i relacji między wyrażeniami tego języka, czyli postulatów znaczeniowych danego jiw, oraz całej swojej wiedzy o języku naturalnym i informacyjno-wyszukiwawczym, w tym ewentualnego podobieństwa znaków obu języków, dokonuje „rekonstrukcji” znaczenia, albo jest jawny i wyraża się za pomocą definicji i eksplikacji<sup>26</sup>, dodatkowo ewentualnie uzupełnianych czy wspieranych relacjami między wyrażeniami jiw w jego systemie paradygmatycznym. Potraktowanie znaczenia jako użycia wymagałoby od projektantów jiw włączenia do procesu projektowania w dużo większym zakresie, niż dziś się to dzieje, badania przykładów używania języka, zarówno przez indeksujących, jak i samych użytkowników, w różnych sytuacjach (kontekstach) społecznych, problemowych, działaniowych itp. W wypadku języków informacyjno-wyszukiwawczych owa rzeczywistość pozajęzykowa charakteryzuje się dwustopniową strukturą, o której pisała Barbara Sosińska<sup>27</sup>. Uzasadniając potrzebę badania znaczeń znaków jiw rozumianych jako ich użycia, mamy na myśli jedynie bezpośrednią płaszczyznę odniesienia języka informacyjno-wyszukiwawczego, tj. rzeczywistość dokumentów opisywanych przez jiw. Znaczeniem znaków jiw byłyby zatem dokumenty źródłowe systemu, w których reprezentacjach znaki te się znalazły. Zauważmy, że struktura semantyczna jiw odwzorowana w jego słowniku stanowi pewną normę językową, podczas gdy

<sup>25</sup> Grzegorzczkova 1995, s. 19.

<sup>26</sup> Dyskusję na temat podobieństw i różnic pomiędzy metodami eksplikacji i definicji, samymi eksplikacjami i definicjami oraz pojęć mających rodziny znaczeń i problemów ich definicji można znaleźć między innymi w: Pawłowski 1986, s. 97-148.

<sup>27</sup> Sosińska 1988.



znaczenia rozumiane jako użycie świadczą o faktycznym obrazie semantyki danego języka informacyjno-wyszukiwawczego. Idealem byłoby pokrywanie się tych dwóch struktur.

Sosińska przyjęła za Ferdynandem de Saussure'em założenie o dualnym charakterze znaku językowego. Na podstawie teorii glossemantycznej Louisa Hjelmsleva wyróżniła w strukturze języka informacyjno-wyszukiwawczego dwa wzajemnie zależne plany: treści i wyrażania, rozszczepiające się z kolei na dwa poziomy: substancji i formy. W opisie kategorii planów jiw wykorzystwała schemat Klasyfikacji Dwukropkowej Ranganathana, wykazującej pewien paralelizm z teoriami de Saussure'a i Hjelmsleva. Przyjęła denotacyjną i konotacyjną<sup>28</sup> interpretację znaczenia, nawiązując do definicji treści i zakresu znaków językowych Ajdukiewicza oraz teorii ekstensji i intensji wyrażen Rudolf Carnapa. W opisie decydującej o interpretacji znaczenia funkcji metainformacyjnej jiw odwołała się do definicji metainformacji Henryka Greniewskiego oraz do definicji językowej funkcji metainformacyjnej Bożenny Bojar. W rezultacie Sosińska wyróżniła:

- dwa typy desygnacji, czyli dwa typy obiektów, które desygnują wyrażenia jiw: desygnaty bezpośrednie – dokumenty opisywane przez język informacyjno-wyszukiwawczy oraz desygnaty pośrednie, czyli obiekty rzeczywistości pozadokumentacyjnej składające się na treść dokumentów;
- dwa typy konotacji: konotację bezpośrednią, którą stanowią zbiory cech desygnatów bezpośrednich, czyli zbiory własności dokumentów oraz konotację pośrednią – własności obiektów rzeczywistości pozadokumentacyjnej.

Istnieje ścisły związek między tymi czterema sferami interpretacji znaczenia znaku jiw. W dokumentacyjnych systemach informacyjno-wyszukiwawczych, żeby uzyskać informację o obiekcie rzeczywistości pozajęzykowej trzeba odwołać się do rzeczywistości językowej (dokumentacyjnej). Z kolei możliwość informowania przez siw o przedmiotach rzeczywistości pozadokumentacyjnej jest uzależniona od uprzedniego odwzorowania w jego zbiorach

---

<sup>28</sup> Denotacyjna i konotacyjna interpretacja *znaczenia* oraz sama opozycja terminologiczna została wprowadzona do nauki przez Milla, według którego wyrażenie denotuje klasę indywidualów, których jest nazwą (tak, iż denotacja jest szczególnym przypadkiem nazywania); jeśli jednak wyrażenie to jest tak zwanym przez Milla konkretnym terminem ogólnym, takim jak *biały* czy *człowiek*, to nie tylko denotuje tę klasę lub jeden z jej elementów, lecz także implikuje jedną lub więcej cech, na mocy których indywiduala są uznawane za elementy tej klasy. Można dostrzec tu związek zachodzący z jednej strony między denotacją a używanym przez innych badaczy wyrażeniem zakres terminu, z drugiej między konotacją a jego treścią. W nowszych pismach filozoficznych Millowskie terminy denotacja i konotacja są często używane na określenie nicco odmiennego przeciwstawienia odniesienia i sensu, wywodzącego się od Fregego. Wyraz konotacja ma z kolei według Milla sugerować, że tak zwana przez niego sygnifikacja atrybutów danego podmiotu jest czymś dodatkowym w stosunku do sygnifikacji (czyli denotacji) wszystkich podmiotów posiadających te atrybuty. U Milla denotacja była relacją łączącą nie wyrażenia w poszczególnych wypadkach ich użycia z ich sygnifikatami, lecz leksemy z całą klasą nazywanych przez nie indywidualów. Na określenie pojęcia pochodnego od pojęcia wypowiedzi i przysługującego nie leksemom jako takim, lecz wyrażeniom w kontekście najczęściej używa się wyrazu odniesienie.

(dokumentach) pewnej wiedzy (językowej i pozajęzykowej) o tych przedmiotach. Projektanci języków informacyjno-wyszukiwawczych najczęściej odwołują się do denotacji bezpośredniej znaków jiw, projektując reguły syntaktyczne jiw oparte na teoriomnogościowych działaniach sumy i iloczynu zbiorów desygnatów bezpośrednich. Do denotacji pośredniej sięga się przy określaniu w systemie leksykalnym jiw relacji mereologicznych oraz tej części relacji kojarzeniowych, która nie odwołuje się do współwystępowania wyrażen w tekstach oraz przy konstruowaniu wyrażen jiw o strukturze semantycznie umotywowanej. O relacjach, w jakie wchodzi wyrażenie jiw, tj. o miejscu wyrażenia w strukturze jiw, decyduje konotacja bezpośrednia i pośrednia, czyli zespół cech semantycznych niezbędnych i wystarczających (tzw. konotacja charakterystyczna) do identyfikacji desygnatów wyrażen wynikających z postulatów znaczeniowych danego jiw. Denotacyjna i konotacyjna interpretacja znaczenia stanowią podstawę semantycznej organizacji pól leksykalnych większości istniejących języków informacyjno-wyszukiwawczych. Ze względu na nasz obecny przedmiot rozważań (znaczenie jako użycie) istotne jest wyróżnienie przez Sosińską dwóch płaszczyzn odniesienia języka informacyjno-wyszukiwawczego – bezpośredniej, czyli rzeczywistości dokumentacyjnej, i pośredniej, czyli rzeczywistości (pozadokumentacyjnej) opisywanej w dokumentach.

Definicje i eksplikacje znaczenia poddane obszernej krytyce przez Ludwiga Wittgensteina w odniesieniu do języków naturalnych<sup>29</sup> nadal odgrywają istotną, jeśli nie decydującą, rolę przy tworzeniu jiw. Przyjęcie koncepcji Wittgensteina nie znaczy, że należałoby zaniechać stosowania definicji i objaśnień. Definicje i eksplikacje są częścią wyjaśniania lingwistycznego. Ale, jak zauważył Wittgenstein, sprawdzają się wówczas, gdy użytkownik języka rozumie generalne formy użycia wyrażenia, tj. gry językowe, i zna aktywności, w kontekście których dane wyrażenie jest używane; tylko wówczas definicje i objaśnienia mogą pomóc odślonić pewne istotne aspekty użycia danego wyrażenia, czyli jego znaczenia. Nie szukając bardziej wyrafinowanego przykładu, wyobraźmy sobie, że kandydatom na studentów bibliologii przedstawimy definicję porządku poziomego ze *Słownika encyklopedycznego terminologii języków i systemów informacyjno-wyszukiwawczych*. Brzmi ona następująco: „Porządek poziomy, porządek cytowania, znormalizowany porządek zapisu – (1) kolejność wyrażen w zdaniach języka informacyjno-wyszukiwawczego (2) zasady określające tę kolejność. Por. formuła fasetowa”<sup>30</sup>. Zapewne większość osób po zapoznaniu się z tą, skądinąd

---

<sup>29</sup> Waga, jaką jesteśmy skłonni przypisywać definicjom i objaśnieniom, jest prawdopodobnie konsekwencją ludzkiej skłonności do generalizowania. Tymczasem nie wszystkie znaczenia językowe dają się sprowadzić do definicji i poklasyfikować, zaś struktura znaczenia nie zawsze daje się rozłożyć do postaci skończonej liczby elementów (ccch) składowych (dystynktywnych). Szczerzej będzie o tym mowa w rozdz. drugim.

<sup>30</sup> Słownik 1993a, s. 118.

poprawnie sformułowaną definicją, nadal nie będzie potrafiła sensownie użyć wyrażenia porządek poziomy w zdaniu lub tekście. Definicje i objaśnienia są rodzajem skrótu chroniącego użytkownika języka przed wyliczaniem wielu przykładów. Można powiedzieć, że rola kontekstu i przykładu nie była do tej pory należycie wykorzystana w teorii i praktyce języków informacyjno-wyszukiwawczych. Projektując jiw, należałoby zatem więcej uwagi zwracać na wykorzystanie przykładów użycia wyrażen w charakterystykach i instrukcjach wyszukiwawczych jako środka prezentowania i posługiwania się znaczeniami w językach informacyjno-wyszukiwawczych. Zauważmy też, że projekty języków informacyjno-wyszukiwawczych na ogół nie są weryfikowane przed wdrożeniem poprzez sprawdzenie na próbie reprezentatywnej, jaka jest ich sprawność jako narzędzi reprezentacji i wyszukiwania dokumentów.

Podkreślając rolę przykładów w uczeniu się i rozumieniu języka informacyjno-wyszukiwawczego i systemu informacyjno-wyszukiwawczego, nawiązujemy do teorii języka Wittgensteina i koncepcji Kuhna dotyczącej opanowania wiedzy naukowej i języka nauki<sup>31</sup>. W większości obecnie funkcjonujących systemów informacyjno-wyszukiwawczych pedagogiczne aspekty dobrych, wyrazistych przykładów użycia wyrażen jiw są wykorzystywane w stopniu i w zakresie, który nie wydaje się zadowolający. Być może przyczyną jest tu brak świadomości, choć nie można też wykluczyć innych czynników. Jednym z nich może być przyjęta koncepcja języka informacyjno-wyszukiwawczego, w tym koncepcja znaczenia, redukująca potrzebę posługiwania się przykładami. Innym – odizolowanie fazy projektowania od fazy implementacji, co, zwłaszcza w wypadku języków budowanych metodą dedukcyjną, może sprawiać, że projektant języka nie jest w stanie wskazać tych przykładów. Naszym zdaniem zarówno użytkownicy, jak i twórcy reprezentacje dokumentów powinni mieć możliwość dostępu do dokumentów prymarnych, w których charakterystykach wyszukiwawczych wystąpiło dane wyrażenie (wyrażenia) jiw. Dokumenty prymarne są bowiem przykładami użycia wyrażen języka informacyjno-wyszukiwawczego. Jest wysoce prawdopodobne, że w przyszłości, być może niedalekiej, pełne teksty dokumentów (elektroniczne kopie dokumentów prymarnych) staną się „wyposażeniem standardowym” każdego siw<sup>32</sup> i każde (lub prawie każde) wyrażenie jiw będzie

---

<sup>31</sup> Kuhn 1968.

<sup>32</sup> Zjawisko to dostrzegają także reprezentanci innych nauk. Jako przykład przytoczymy fragment referatu wygłoszonego przez Ryszarda Tadcusiwicza na konferencji *Współpraca bibliotek naukowych w zakresie automatyzacji*, Kraków 16-19.11.1998. „Jednym z pierwszych objawów zmierzania w kierunku społeczeństwa informacyjnego będzie to, że już wkrótce nieuchronnie dojdzie (w skali ogólnoświatowej!) do lokowania na serwerach internetowych różnych publikacji (czasopism fachowych, gazet codziennych, książek itp.), a także stosownych informacji katalogowych. Istotna różnica w stosunku do stanu obecnego, także charakteryzującego się przebogatą ofertą różnych internetowo dostępnych danych i wiadomości, polegać będzie na tym, że obecnie gromadzi się i udostępnia głównie różnego rodzaju skróty, streszczenia i omówienia, podczas gdy

mogło mieć w niej dobry, wyrazisty przykład, lub lepiej przykłady użycia. Zauważmy, że taka pełnotekstowa baza przykładów byłaby też, zwłaszcza dla nowych indeksatorów, cenną pomocą w uczeniu się jiw. Innym potencjalnym obszarem wykorzystania przykładów jest udostępnianie użytkownikom zapisów przebiegu wybranych realnych, nie hipotetycznych, wyszukiwań. Wprawdzie wiele wyszukiwań jest unikatowych i trudno porównywalnych, ale przykłady można byłoby traktować jako narzędzie pedagogiczne, ilustrujące pewne regularności zachodzące w systemie i w języku – to, jak może wyglądać dobre wyszukiwanie i jak posługiwać się poszczególnymi wyrażeniami jiw, jak budować z nich zdania i teksty. Tradycyjnie wiele systemów informacyjno-wyszukiwawczych rozwiązuje te problemy zapewniając użytkownikowi pomoc pośrednika, wykwalifikowanego bibliotekarza czy pracownika informacji, który zna system, jiw, techniki i strategie wyszukiwawcze. Pośrednicy często mają nawet głęboką wiedzę o systemie, ale na ogół mniejszą wiedzę o przedmiocie wyszukiwania, a na pewno nie mają tej świadomości potrzeb i oczekiwań wobec systemu, którą ma użytkownik. Użytkownik z kolei często nie potrafi zwerbalizować swoich potrzeb. Dzieje się tak na przykład z powodu „niewystarczającego stanu wiedzy” (ang. *anomalous state of knowledge*, ASK), o którym obszernie pisał Nicholas Belkin<sup>33</sup> i inni autorzy podzielający to założenie, choć różnie je nazywający, interpretujący i komentujący. Niewystarczający stan wiedzy może też odnosić się do umiejętności i rozumienia tego, jak wygląda wyszukiwanie w danym systemie i jak wyrażenia jiw są używane do reprezentowania i wyszukiwania dokumentów. Użytkownicy, zwłaszcza nowicjusze, często tej wiedzy nie posiadają.

Użytkownik formułuje zapytanie odwołujące się do cech treściowych dokumentów na ogół wówczas, gdy nie zna żadnych informacji kontekstowych identyfikujących relewantne dokumenty. Wyszukiwanie według kryteriów treściowych odbywa się w różnorodnych sytuacjach, zwanych w literaturze sytuacjami problemowymi, współtworzonych przez różne rodzaje działań, aktywności czy zadań motywujących do podjęcia poszukiwań informacji. Na ogół sytuacje te charakteryzują się własnym językiem, a zwłaszcza słownictwem. Ten specjalistyczny, żargonowy czy gwarowy język sprawia, że takie a nie inne słowa, w taki a nie inny sposób są używane do określenia przedmiotu zainteresowania. Ponadto istota pewnych działań, takich jak na przykład

---

już wkrótce elektronicznej dystrybucji podlegać będą zasoby w postaci pełnych tekstów odpowiednich dzieł. Aktualnie się tego nie robi ze względu na ograniczenia techniczne i ekonomiczne napotykanie zarówno w zakresie gromadzenia dużych ilości informacji na serwerach (duży zasób pamięci i obsługujący ją procesor drogo kosztują) oraz ze względu na trudności w procesie dystrybucji (przesyłanie dużych zbiorów informacji pomiędzy serwerem i odbiorcą poważnie obciąża finansowo obie strony). Jednak postęp techniczny wymusza stale radykalną obniżkę kosztów i już wkrótce dzisiejsze ograniczenia będą równie anachroniczne, jak dla nas obecnie limity mocy obliczeniowych sławnych komputerów RIAD (...). Tadcusiewicz 1998, s. 10.

<sup>33</sup> Belkin *passim*.

przygotowywanie linii obrony w procesie sądowym, sprawdzanie, czy dany wynalazek jest objęty ochroną patentową, przygotowywanie się do podjęcia badań naukowych, wymaga, aby wyszukiwanie spełniało warunek kompletności całkowitej lub najwyższej możliwej do osiągnięcia. Tymczasem języki informacyjno-wyszukiwawcze najczęściej ani nie uwzględniają „żywego” słownictwa użytkowników, które mogłoby być zidentyfikowane w wyniku analizy zachowań językowych użytkowników, ani nie zawsze gwarantują środki osiągnięcia pożądanej kompletności.

Praktyka pokazuje też, że różni użytkownicy w różnych sytuacjach problemowych często pytają o te same dokumenty w różny sposób. Różnorodność ta objawia się zarówno na poziomie wyrażania, jak i na poziomie treści zapytań i instrukcji wyszukiwawczych. Tradycyjnie systemy informacyjno-wyszukiwawcze nie były wrażliwe na to zjawisko. Idiosynkrazje w sposobach, w jakie użytkownicy mogą pytać o dokumenty, były i często nadal są traktowane jako aberracje. Postawa ta znajduje swoją ekstremalną postać w skrajnie normatywnym podejściu do indeksowania, kiedy to decyzję indeksującego traktuje się jak wyrocznię co do „poprawności” lub „niepoprawności” danej reprezentacji dokumentu. W konsekwencji rola użytkownika nierzadko polega na podejmowaniu prób odgadnięcia, jakie mogą być charakterystyki wyszukiwawcze relewantnych dla niego dokumentów. Jak powiedziałby Wittgenstein, przyczyną jest tu mylenie dwóch różnych poziomów i form tworzenia reprezentacji informacji – opisywania dokumentu, a właściwie jego treści, z innymi formami reprezentacji czy klasyfikacji. Istnieją podstawy do przypuszczania, że zarówno teoretycy, jak i praktycy żyw w pewnych, nie tak rzadkich przypadkach, nie czynią rozróżnienia pomiędzy opisem i klasyfikacją w nauce a opisywaniem i klasyfikowaniem dokumentów. Używając terminów wypracowanych przez Sosińską, można powiedzieć, że mylą lub łączą płaszczyznę denotacji i konotacji bezpośredniej z płaszczyzną denotacji i konotacji pośredniej. Języki informacyjno-wyszukiwawcze od początku swego istnienia, czyli od pierwszych klasyfikacji piśmiennictwa, pozostawały w związku z klasyfikacjami nauk i ich obiektów. Związek ten obserwuje się i we współcześnie tworzonych i wykorzystywanych żyw. Nie stanowi on bynajmniej o słabości języka, przeciwnie, często jest źródłem jego stabilności, nie może być jednak jedyną czy centralną jego zasadą. Posłużymy się tu niezbyt wyrafinowanym, ale za to wyrazistym przykładem języka informacyjno-wyszukiwawczego dla biblioteki publicznej, nie pełniącej funkcji biblioteki naukowej. Obecnie w Polsce w bibliotekach tego typu w funkcji żyw wykorzystuje się najczęściej albo Uniwersalną Klasyfikację Dziesiątą, albo język haseł przedmiotowych Biblioteki Narodowej. Oba języki wyraźnie nawiązują do naukowego obrazu świata i klasyfikacji nauk. Tymczasem użytkownik biblioteki publicznej poszukujący tam informacji najczęściej operuje innym obrazem świata. Opartym nie na taksonomiach naukowych, a raczej na potocznym lub „popularnonaukowym” układzie kategoryalnym. Najczęściej również dokumenty

zgrupowane w tych bibliotekach nie są pracami naukowymi i zdają się być „na siłę” wtłoczone w ramy systematyki naukowej<sup>34</sup>. Jeśli zgodzimy się, że jedną z dwóch głównych funkcji języków informacyjno-wyszukiwawczych jest funkcja informacyjna w stosunku do rzeczywistości dokumentacyjnej, to w budowaniu jiw należałoby uwagę zwracać na obraz struktury informacji, jaki wyłania się z dokumentów, a następnie struktury informacyjnej pośrednio wskazywanej rzeczywistości, w tym systemu nauki. Chcemy się zastrzec, że naszym celem nie jest sformułowanie alternatywy, lecz raczej warunku komplementarności. Wyrażony wcześniej postulat nie jest niczym nowym w teorii jiw, ale jego implikacje praktyczne są raczej mierne ze względu na relatywnie dużą trudność stworzenia instrumentów projektowych, które pozwoliłyby na jego praktyczną realizację.

Sosińska w cytowanej pracy doktorskiej stwierdziła, że w językach informacyjno-wyszukiwawczych „problem odwzorowania naiwnego i naukowego obrazu świata nie istnieje, języki te bowiem, jako przeznaczone do opisu literatury specjalistycznej, zmierzają do odwzorowania struktury terminologii poszczególnych dziedzin nauki”<sup>35</sup>. Nasze zdanie w tej kwestii jest odmienne, o czym świadczy samo zajęcie się problematyką kategoryzacji w jiw. Problem odwzorowania ogólnego<sup>36</sup> i naukowego obrazu świata nie tylko istnieje, ale

---

<sup>34</sup> W informacji naukowej tym, którzy potrzebują dostępu do wiedzy eksperckiej, poświęcano przez lata sporo uwagi. Można zaryzykować stwierdzenie, że badania zachowań w zakresie poszukiwania informacji (zapotrzebowania na informację) dałyby się w większości sprowadzić do badań zachowań naukowca poszukującego informacji. Można spierać się, czy poświęcanie uwagi naukowcom przez całe lata było dobrą czy złą przysługą dla informacji naukowej, zwłaszcza jeśli chodzi o jej behawioralne aspekty. Postępowanie w nauce jest kwestią wysoce podporządkowaną wzorcom wraz ze specyficznymi normami określającymi zachowania dopuszczalne (akceptowalne), zaś rola poszukiwania informacji i jej wykorzystania jest w tych normach raczej jasno zdefiniowana. Naukowcy mają do czynienia z problemami stosunkowo dobrze ustrukturalizowanymi, dla których mogą już istnieć drogi dochodzenia do rozwiązania (wzorccc), a wiele innych osób mogło tymi drogami wcześniej podążać przez całe lata. Analizowanie tych wcześniejszych prac, poszukiwanie pominiętych skojarzeń, przyjmowanie podejść wcześniej uważanych za ślepe uliczki, podejmowanie obiecujących wzmianek przeczonych przez wcześniejszych badaczy jest zasadniczą częścią pracy naukowca. Wyrażanie uznania dla wcześniejszych prac poprzez cytowanie, a przez to odnotowywanie użycia dokumentu również jest częścią reguł gry, lub inaczej mówiąc, częścią systemu relewancji i wzorców naukowca. W konsekwencji, identyfikacja informacji potencjalnie użytecznej, projektowanie systemów informacji oraz dostarczanie usług informacyjnych jest dla profesjonalisty w dziedzinie informacji naukowej stosunkowo prostą sprawą bynajmniej nie z powodu jego systemu relewancji, bo ten wiąże się z dokumentami i ich wykorzystaniem, ale dlatego, że pokrywa się z systemem relewancji naukowca – podobną wagę przykładają do dokumentów, wyszukiwania oraz przekazywania informacji. Jednakże nauka jest tylko jedną ze sfer życia człowieka. Inne rodzaje pracy, w jaką angażują się ludzie, są zbyt liczne by je wymieniać, lecz każda jest odmienna i niewielu z nich przysługują te same cechy co nauce. Traktowanie wszystkich potencjalnych użytkowników informacji tak, jakby byli naukowcami i patrzenie na wszystkie konteksty pracy jak na kontekst w nauce jest, powiedzmy, co najmniej nieuprawnione.

<sup>35</sup> Sosińska 1988, s. 149.

<sup>36</sup> Obraz ogólny, język ogólny bywają też nazywane potocznymi lub naiwnymi.

jest jednym z tych, które do tej pory albo nie były zauważane albo – z różnych powodów zredukowane do problematyki obrazu naukowego. A przecież nie wszystkie dokumenty gromadzone w bibliotekach i charakteryzowane za pomocą jiw odwzorowują naukowy obraz świata oraz nie zawsze, a w pewnych typach bibliotek na ogół, pytania formułowane przez użytkowników nie nawiązują do obrazu naukowego a właśnie do mniej lub bardziej zindywidualizowanego obrazu potocznego. Ignorowanie tej sytuacji nie wydaje się ani uprawnione, ani usprawiedliwione.

W językoznawstwie panuje raczej powszechna zgoda co do tego, że znaczenia znaków językowych są efektem dwojakiego uwarunkowania. Z jednej strony są to realia świata opisywanego przez język<sup>37</sup>, z drugiej zaś – interpretacje tej rzeczywistości. Interpretacja jako działanie subiektywne zakłada wielość możliwych spojrzeń na otaczający świat. W konsekwencji prowadzi to zarówno do odmiennego semantyczno-leksykalnego ukształtowania różnych języków, jak też do powstawania różnych odmian semantyczno-leksykalnych w obrębie tego samego języka etnicznego, a nawet w obrębie języka nauki. Obrazy świata tworzone przez język nauki i język ogólny są tylko przykładem możliwego zróżnicowania wewnątrzjęzykowego. Obraz naukowy kształtowany jest przede wszystkim przez odniesienia do rzeczywistości, dążąc do możliwie wiernego jej odwzorowania, do minimum redukując subiektywizm człowieka. Obraz świata zawarty w języku ogólnym uwzględnia zarówno obiektywnie dane cechy rzeczywistości, jak i inne uwarunkowania rządzące myśleniem i ocenami człowieka. Naukowe poznanie świata i pochodny od niego naukowy obraz świata w dużej mierze, choć nie całkowicie, są wspólne ludziom mówiącym różnymi językami, z kolei potoczny obraz świata jest utrwalony i odtwarzany poprzez język ogólny i dlatego każdy język etniczny ma sobie właściwy obraz świata, a właściwie obrazy świata. Na przykład, w języku ogólnym specjalistycznemu terminowi *motyl* z taksonomii zoologicznej (entomologicznej) odpowiadają dwa słowa: *motyl* – najbliższy odpowiednik entomologicznego terminu *motyl dzienny* i *ćma* – ogólny odpowiednik terminu *motyl nocny*. U podstaw binarnego podziału tej grupy leżą nie tylko obiektywne właściwości świata owadów, co widoczne jest w taksonomii entomologicznej (*motyl dzienny* – *motyl nocny*), ale także ogólniejszy system kulturowych preferencji człowieka widocznych w tej grupie leksykalnej i rozciągających się również na inne sfery leksyki. W języku ogólnym nazwa *motyl* konotuje między innymi lekkość, radość i urodę, *ćma* zaś, na skutek etymologicznego związku z ciemnością, nie ma takich konotacji.

---

<sup>37</sup> Jednym ze zwolenników tezy, że znaczenia znaków językowych są determinowane własnościami opisywanych obiektów był Witold Doroszcwski. „(...) wymienianie pierwiastków treści znaczeniowej słów w definicji hasel jest jednoznaczne z wymienianiem cech odpowiadających słowom pozajęzykowych desygnatów”. Doroszcwski 1982, s. 50.

Z ogólnych własności języków naturalnych wynika, że znaczenie znaku językowego i jego użycie pozostają pod wpływem zjawiska nieograniczonej semiozy<sup>38</sup>, wyrażającej się w tym, że „nie ma praktycznego lub teoretycznego limitu liczby różnych sposobów (słowa, frazy, zdania), jakimi ta sama idea czy znaczenie mogą być wyrażone”<sup>39</sup>. Jak pisał Adam Heinz „funkcjonowanie jako znak (*semiosis*) jest wypadkową trzech czynników: 1) rzeczy oznaczającej (*sign-vehicle*), 2) rzeczy oznaczanej (*designatum, denotatum*) oraz 3) człowieka (*interpretant*); do tego dochodzą dwa dalsze: 4) proces interpretacji, czyli dyspozycja do odpowiedniego zachowania się interpretanta, oraz 5) kontekst sytuacyjny”<sup>40</sup>. Zjawisko nieograniczonej semiozy ma ogromne znaczenie dla opracowania i wyszukiwania informacji, uświadamia bowiem wagę i konieczność zapewnienia w języku informacyjno-wyszukiwawczym środków do panowania nad tym, niekiedy uciążliwym dla projektanta jiw, bogactwem systemów znaków językowych.

Zdaniem Blaira<sup>41</sup> problem nieograniczonej semiozy w systemach informacyjno-wyszukiwawczych należałoby widzieć w dwojakiej perspektywie. Po pierwsze, istnieje (przynajmniej teoretycznie) ogromna liczba różnych (nie tych samych) dokumentów, które mogą być reprezentowane przez tę samą charakterystykę. Po drugie, istnieje duża liczba charakterystyk treściowych, które mogą być sensownie zastosowane do jednego dokumentu. Tradycyjnie teoria jiw i indeksowania, choć świadoma rozbieżności, niejasności i niespójności ujawniających się w tworzeniu charakterystyk i instrukcji wyszukiwawczych, preferowała myślenie, że zjawiska te są raczej incydentalnymi czy nieistotnymi zakłóceniami w sprawnie funkcjonującym modelu. Brało się to między innymi z mylenia, o czym wcześniej wspominaliśmy, dwóch płaszczyzn informacji – płaszczyzny informacji odwzorowanych w dokumentach ze wszystkimi jej możliwymi błędami i idiosynkrazjami oraz płaszczyzny informacji w rzeczywistości, w tym w systemie nauki. Negatywną rolę odegrało tu także naiwne i błędne przekonanie, że treści dokumentów są na tyle proste, że mogą być efektywnie i wyczerpująco reprezentowane za pomocą pewnej liczby wyrażen jiw o dobrze zdefiniowanych znaczeniach. Tymczasem, nawiązując do koncepcji Wittgensteina, należałoby przypomnieć, że wyrażenia, zarówno języków naturalnych, jak i jiw, mają użycia i że na te użycia powinniśmy patrzeć, a nie (tylko) na definicje. Użycia są trudne do ujęcia w ramach definicji czy schematów definicyjnych, a w wielu wypadkach mogą w ogóle nie poddawać się definiowaniu – można jedynie je pokazać. A zatem możliwy jest rozdzźwięk między zaprojektowanym systemem znaczeń języka informacyjno-wyszukiwawczego a systemem jego realnych użyć.

<sup>38</sup> Semioza, według Peirce’a i Morrisa, użycie czegoś jako znaku. Pele 1984, s. 203.

<sup>39</sup> Blair 1990, s. 169.

<sup>40</sup> Heinz 1983, s. 435.

<sup>41</sup> Blair 1990, s. 169-171.



Wskazując ważniejsze problemy tworzenia reprezentacji dokumentów, nie można nie wymienić zagadnienia tzw. adekwatności<sup>42</sup> charakterystyki wyszukiwawczej. Jednym z nielicznych autorów, którzy zajmowali się tym problemem był M. E. Maron<sup>43</sup>. Zdaniem Marona jedynym kryterium oceny, czy treść dokumentu jest właściwie reprezentowana przez funkcjonującą w systemie charakterystykę wyszukiwawczą, jest sprawdzenie, czy użytkownik, który uznał dany dokument za relewantny, pytał w taki sposób, w jaki została sformułowana charakterystyka. Zauważmy, że przy takim podejściu do problemu reprezentacji informacji w systemie unikamy pułapki przesadnego normatywizmu w indeksowaniu, przenosząc właściwy ciężar uwagi na językowe preferencje użytkownika.

Maron zaproponował koncepcję trzech aspektów postrzegania treści dokumentu i co za tym idzie trzech znaczeń *o czym jest* (ang. *meaning of about*). Pierwszy nazwał S-about (Subjective-about). Jest to wewnętrzny, subiektywny aspekt odbiorcy informacji. Ten punkt widzenia (aspekt) jest tym, który odpowiada najczęstszemu, potocznemu, zwykłemu rozumieniu pojęcia tematu. W myśl tej interpretacji to, o czym jest dokument  $D_i$  musi mieć związek z osobistym i subiektywnym doświadczeniem odbiorcy, będącym rezultatem przeczytania  $D_i$ . S-about jest bytem mentalnym, psychicznym i jako taki w zasadzie nie może być analizowany w kategoriach obiektywnych. Z subiektywnym S-about wiąże się obiektywne lub behawioralne *o czym jest*. Maron nazwał je O-about (Objective about). Interpretacja O-about przeciwstawia to, co zewnętrzne i obserwowalne, temu, co wewnętrzne i subiektywne. O-about odnosi się do aktualnego lub potencjalnego zachowania osoby pytającej lub szukającej dokumentu; jest zewnętrznym korelatem S-about, może bowiem być obserwowane w sytuacji wyszukiwawczej. Zatem to, o czym dokument  $D_i$  jest obiektywnie dla osoby  $X$  wskazuje termin (lub terminy), za pomocą których  $X$  pytałby (lub szukał), jeśli potrzebowałby informacji, jaką zawiera  $D_i$ . Trzeci aspekt to R-about (Retrieval-about). R-about, podobnie jak O-about, jest obiektywne i behawioralne. W przeciwieństwie do O-about odnosi się jednak nie do zachowania pojedynczej osoby, ale klasy osób – tych, którzy tworzą populację użytkowników danej biblioteki czy systemu informacyjno-wyszukiwawczego. R-about jest prezentowane przez rozkład prawdopodobieństwa we wszystkich możliwych kategoriach informacji (nazywanych przez wszystkie dostępne wyrażenia  $I_j$ ), gdzie dla każdego wyrażenia  $I_j$  prawdopodobieństwo, o którym mowa, jest prawdopodobieństwem tego, że jeśli dana osoba uznała dokument  $D_i$  za relewantny, to  $I_j$  jest terminem (jednym z terminów), którego ta osoba użyła do sformułowania zapytania informacyjnego. Oczywiście na uznanie dokumentu za relewantny ma wpływ wiele czynników, wśród których treść dokumentu jest elementem kluczowym, ale nie wyłącznym. Inne to

---

<sup>42</sup> Pewne spostrzeżenia dotyczące adekwatności charakterystyki wyszukiwawczej można znaleźć w artykule Woźniak 1998.

<sup>43</sup> Maron 1977.

rozumiałość, wiarygodność, ważność, aktualność, forma prezentacji informacji itd.

Pewnym przyczynkiem do zagadnienia reprezentacji są też badania szerokości i szczegółowości indeksowania przeprowadzone przez Dona Swansona<sup>44</sup>. Eksperyment Swansona polegał na sformułowaniu jak największej liczby sensownych pytań, na które odpowiedzi można było znaleźć w poszczególnych badanych dokumentach, a następnie na zaindeksowaniu tych pytań. Okazało się, że bez większego trudu dało się sformułować więcej niż 300 pytań do stosunkowo krótkich, liczących kilka tysięcy słów, artykułów. Tymczasem w bibliotekach szerokość i szczegółowość indeksowania zwykle są określane intuicyjnie przez osoby zatrudnione przy indeksowaniu. Nie ma dowodów, że intuicja ta jest w jakikolwiek sposób skorelowana z mierzalnymi parametrami efektywności wyszukiwania, choćby takimi jak kompletność i dokładność.

Indeksujący czy osoby pomagające użytkownikom w wyszukiwaniu informacji nierzadko określają ich zapytania jako nieprecyzyjne, niejasne czy wieloznaczne. Tymczasem często jest to pozorna nieprecyzyjność, niejasność czy wieloznaczność wynikająca na przykład z ignorowania wagi informacji kategoryzacyjnej, z „oderwania” pytania od jego kontekstu. Świadczyłyby o tym fakt, że przy szukaniu informacji kanałami nieformalnymi, dla których charakterystyczna jest większa interakcyjność i sprzężenie zwrotne, zarzut nieprecyzyjności na ogół się nie pojawia. Blair<sup>45</sup> charakteryzuje informacje treściowe i kontekstowe za pomocą metafory, której swobodną parafrazę zamieszczamy poniżej.

Używanie w codziennym życiu lub w małych systemach informacyjno-wyszukiwawczych, gdzie użytkownikowi często towarzyszy wykwalifikowany pracownik, określeń odwołujących się do informacji treściowej przypomina opisywanie nieznanego komuś, kto musi spotkać tę osobę na lotnisku. Oprócz informacji opisujących samą osobę istotne będą także informacje kontekstowe, takie jak: na które lotnisko przylatuje, kiedy, którym rejssem itd. Na ogół informacje treściowe opisujące osobę, na przykład „mężczyzna średniego wzrostu, koło czterdziestki, szpakowaty z brodą i wąsami”, pozwalają wyeliminować z grupy pasażerów danego lotu większość tych, którzy nie spełniają warunków opisu. Na ogół też więcej niż jedna osoba, czasem dwie, czasem kilka, odpowiada posiadanemu opisowi i wówczas, jeśli na podstawie posiadanych informacji nie możemy dokonać ostatecznego wyboru, najprościej jest spytać te osoby, czy są poszukiwanym Iksem. Zwykle język naturalny ma wystarczające środki do realizacji tego typu zadań, choć oczywiście niektórzy użytkownicy języka potrafią lepiej od innych sporządzić i wykorzystać taki opis, a z kolei pewni ludzie posiadają więcej cech „dystyngtywnych” ułatwiających ich opis i odróżnienie od innych.

---

<sup>44</sup> Swanson D. R.: *Studies of indexing depth and retrieval effectiveness*. Unpublished report National Science Foundation Grant GN 380, February 1966. Cyt. za: Blair 1990, s. 170.

<sup>45</sup> Blair 1990, s. 172-173.

Wyobraźmy sobie teraz, że mamy opisać nieznaną osobę komuś, kto ma ją spotkać na stadionie, na którym zgromadził się trzydziesto- czy czterdziestotysięczny tłum. Opis szukanej osoby, jak przytoczony wcześniej czy nawet dokładniejszy, daje nikłą szansę sukcesu. Sytuacja ta przypomina sytuację użytkownika szukającego informacji w dużym systemie informacyjno-wyszukiwawczym. Chcąc odnaleźć tę osobę należałoby posłużyć się nie tylko o wiele jej dokładniejszym opisem werbalnym, ale i większą liczbą kategorii informacji kontekstowych, takimi jak: to, w którym sektorze zwykle siedzi; że na ogół nosi kurtkę w barwach swojej drużyny; że przychodzi ze swoim ośmioletnim synem itd.

W pewnych sytuacjach, gdy szczególnie zależy nam na sukcesie, należy dodatkowo posłużyć się środkami pozajęzykowymi. Wyobraźmy sobie poszukiwanie przestępcy ukrywającego się w dużym mieście. Na pewno nieodzowna jest wówczas fotografia szukanej osoby, najlepiej jeśli są to dwie fotografie: en face i z profilu, przy czym na fotografii en face osoba powinna patrzeć w obiektyw i mieć „neutralny” wyraz twarzy. Fotografii powinien towarzyszyć zestaw informacji kontekstowych, między innymi takich jak: wzrost, waga, kolor włosów i oczu, znaki szczególne, umiejętności, zawód, kontakty itd. Aby szukanie danej osoby było działaniem sensownym, potrzebne są zatem: bardzo dokładny opis werbalny tej osoby, czyli informacje treściowe, dobre przykłady (tutaj fotografie pokazujące, jak wygląda, czy może wyglądać, szukana osoba), jak najwięcej informacji kontekstowych i stałe inteligentne sprzężenie zwrotne.

Kontekst jest nadal niedocenianym czynnikiem w tworzeniu reprezentacji dokumentów i w wyszukiwaniu. Prawdopodobnie jest to powodowane tym, że wiele nawet komercyjnych systemów informacyjno-wyszukiwawczych, nie mówiąc o zautomatyzowanych katalogach bibliotecznych, jest budowanych w nawiązaniu do tradycyjnego modelu biblioteki i reprezentacji dokumentu. Ani ogromny wzrost liczby wydawanych i gromadzonych dokumentów, ani rewolucja komputerowa nie wpłynęły istotnie na zmianę podejścia do problemów reprezentacji i wyszukiwania informacji. Użytkownik wydaje się być jedynie werbalnie punktem odniesienia dla wielu działań, także projektantów jiw. Sformułowanie to nie jest aktem oceny, a jedynie stwierdzeniem faktu i co za tym idzie zawiera supozycję potrzeby zmiany. Znowu powołamy się na Wittgensteina, który podstawę aktywności językowych widział w faktach pozajęzykowych. Wynikałoby z tego, że informacje dotyczące przynajmniej typowych działań, w wyniku których powstają lub w trakcie których wykorzystywane są dokumenty związane z określonym tematem, mogą być użyteczne dla reprezentowania tych dokumentów w celu ich późniejszego odszukania. Informacje te najczęściej przybierają postać informacji kontekstowych, o których wcześniej była mowa. Z tezy Wittgensteina wynika też jeszcze jeden wniosek dla projektantów języków informacyjno-wyszukiwawczych. Informacje kontekstowe, o które się tu dopominamy, nie zawsze muszą

być jawnie zawarte w relewantnym dokumencie. Czasami można je zidentyfikować (wydedukować) jedynie znając sytuacje, w których dokument (zawarte w nim informacje) znajdują właściwe zastosowanie. Zagadnienie kontekstu<sup>46</sup> w systemach i językach informacyjno-wyszukiwawczych należałoby zatem widzieć w dwóch wzajemnie się przenikających perspektywach – kontekstu użycia języka w celu wskazania, opisanie rzeczywistości pozajęzykowej i jej elementów oraz kontekstu użycia języka w celu „czynienia rzeczy”<sup>47</sup> w rozumieniu na przykład gier językowych czy aktów mowy i illokucji<sup>48</sup>.

W dotychczasowej teorii i praktyce języków informacyjno-wyszukiwawczych najwięcej uwagi (jeśli nie całą) poświęcano różnym problemom opisu treści. Zjawisko to znalazło wyraz między innymi w dość powszechnie akceptowanym stwierdzeniu, że jiw nie mają pragmatyki. Jeśli jednak pragmatyka „zajmuje się badaniem performancji językowej”<sup>49</sup> będącej „aktualizacją (przejawem) kompetencji językowej w konkretnym komunikacie”<sup>50</sup>, to pragmatyka jiw zajmowałaby się przejawami tworzenia i rozpoznawania tekstów jiw. Większe zwrócenie się ku pragmatycznym aspektom jiw musiałoby pozostawać w ścisłym związku z analogicznymi tendencjami w językoznawstwie, bez którego dorobku metodologicznego i przedmiotowego nie jest chyba możliwy właściwy rozwój teorii i praktyki jiw, będącej ważną i niezbywalną częścią nauki o informacji. Do tej pory pragmatyka jeśli była przedmiotem badania, to zwykle w ograniczeniu tylko do jednej kategorii użytkowników języków informacyjno-wyszukiwawczych – osób indeksujących<sup>51</sup>. Dziś wydaje się, że rozszerzenie możliwości jiw tkwi w tej sferze języka – w posługiwaniu się jiw w procesach komunikacji między szukającym i tworzącym reprezentacje dokumentów w systemie. Potencjalne badania widzielibyśmy

---

<sup>46</sup> Odwoływanie się do informacji kontekstowej jest tylko jedną z prób wskazywania możliwych sposobów doskonalenia procesów informacyjno-wyszukiwawczych. W literaturze można znaleźć inne propozycje. Jedną z nich jest koncepcja przedstawiona w artykule: O'Connor 1993.

<sup>47</sup> Nawiązuje się tu do pracy J. L. Austina *How to Do Things with Words*, co w przekładzie Grzegorzycykowej brzmi „Jak czynić rzeczy za pomocą słów”. Grzegorzycykowa 1995, s. 84.

<sup>48</sup> Początki teorii aktów mowy wiążą się z refleksją filozoficzną nad językiem, sformułowaną przez brytyjskich filozofów języka, należących do tzw. oksfordzkiej szkoły analitycznej, przede wszystkim Austina. Austin zaobserwował, że pewne wypowiedzi o postaci konstatacji zachowują się niejednakowo, jeśli chodzi o cechę prawdziwości/fałszywości, mianowicie niektóre z nich nie podlegają ocenie prawdziwościowej. Nazwał je performatywami. Wypowiedzi performatywne (wykonawcze), tj. stwarzające pewną rzeczywistość, charakteryzują się tym, że można je oceniać najwyżej jako udane lub nieudane (fortunne/niefortunne). Austin pogłębił pojęcie aktu mowy, a przede wszystkim stworzył tzw. teorię illokucji. Zdaniem Austina każda wypowiedź ma swój aspekt illokucyjny, określający intencję mówienia. Do najważniejszych kontynuatorów i krytyków koncepcji Austina należą między innymi Searle, Strawson i Grice.

<sup>49</sup> Słownik 1993a, s. 119.

<sup>50</sup> Ibidem s. 113.

<sup>51</sup> Określenia *indeksujący* (*indeksator*) używamy jako nazwy kategoryjnej dla wszystkich osób zajmujących się tworzeniem reprezentacji dokumentów oraz personifikacji reguł indeksowania automatycznego.

jako projekt interdyscyplinarny łączący między innymi nauki (lub ich działy) mające za przedmiot użytkownika informacji, język, komunikację, relewancję i samą informację. Jednym z założeń tych badań byłoby zapewne odrzucenie tradycyjnego w nauce o informacji pojmowania relewancji<sup>52</sup> jako czegoś, co jest dane w systemie informacyjno-wyszukiwawczym i co należy jedynie odkryć. Podejście takie jest bowiem charakterystyczne tylko dla części sytuacji wyszukiwawczych, mianowicie tych, które w literaturze określa się mianem dydaktycznych, reprodukujących istniejącą wiedzę.

Drugą sferą nauki, w której widzimy potencjalne możliwości nowego wpływania na teorię i praktykę informacyjno-wyszukiwawczą, są nauki kognitywne<sup>53</sup>. Nauki kognitywne są stosunkowo młodą dyscypliną, do której chętnie się „przyznają” różni badacze, podczas gdy inni uprawiając naukę w duchu kognitywnym unikają tego określenia. Ma też swój dorobek językoznawstwo kognitywne, którego ważniejsze z naszego punktu widzenia konstatacje zreferujemy w rozdziale drugim. Obecnie zajmujemy się, co uzasadnialiśmy wcześniej, schematycznym określeniem zakresu i specyfiki badań kognitywnych w nauce o informacji, na koniec przedstawiając prywatne systemy informacji historyków amerykańskich jako przykład podejścia kognitywnego do kategoryzowania i wyszukiwania informacji.

## KOGNITYWNE ASPEKTY NAUK

W filozofii modele kognitywne oznaczają opozycję wobec klasycznej arystotelesowskiej teorii poznania i obiektywizmu jako jej naczelnego założenia, przyjmowanego za podstawę na przykład tradycyjnej psychologii i językoznawstwa strukturalnego. Opozycja ta wyraża się w zakwestionowaniu teorii kategoryzacji Arystotelesa, w myśl której świat składa się z bytów o ustalonych cechach, które to byty pozostają między sobą w pewnych relacjach, obiektywnych i niezależnych od umysłu człowieka. Równocześnie kwestionowana jest jej logiczna podstawa – teoria mnogości.

W psychologii kognitywizm zakłada, iż modelowanie reprezentacji wiedzy i przetwarzania informacji przez człowieka polega na wykorzystaniu struktur języka naturalnego do badania procesów poznawczych i percepcji. Charakterystyczne dla psychologii kognitywnej jest traktowanie mózgu ludzkiego jako procesora informacji, w którym wyróżnia się kanały kodowania, prze-

---

<sup>52</sup> Relewancji poświęcona jest między innymi rozprawa doktorska Elżbiety Artowicz i książka opublikowana na jej podstawie. Artowicz 1997, 1990. Tam też można znaleźć obszerne zestawienie literatury na temat relewancji.

<sup>53</sup> W literaturze polskiej można zaobserwować, że część autorów posługuje się terminem *nauki kognitywistyczne* (np. Tabakowska 1995), część zaś *nauki kognitywne* (np. Łozowski 1994). W pracy używamy określenia *kognitywne*.

chowowania i przekazu informacji oraz próbuje określić formalne własności procesów, jakim podlega informacja. Peter Lindsay i Donald Norman stwierdzają: „gdy usiłujemy znaleźć zasady organizacji mózgu ludzkiego, celowe jest przeanalizowanie zasad organizacji i struktur wszelkich systemów informacyjnych. Nacisk kładziemy tu na słowo zasady. Szczegóły dotyczące różnych sposobów funkcjonowania oraz mechanizmy uruchamiania tych systemów w tym miejscu zupełnie nas nie interesują. Jeśli jesteśmy w stanie stwierdzić, że dany system posługuje się metodą heurystyczną analizy środków-celów, to nie ma znaczenia, czy zbudowany jest on z neuronów, elektronicznych obwodów scalonych, czy też przekładni i dźwigni – heurystyka we wszystkich wypadkach jest taka sama”<sup>54</sup>. Akcentuje się więc dynamizm i proceduralny charakter reprezentacji rzeczywistości w mózgu człowieka oraz podleganie ciągłym transformacjom na skutek zmiany stanów pamięci.

W aspekcie językoznawczym<sup>55</sup> podejście kognitywne oznacza próbę sformułowania nowej filozofii języka, traktowanej jako opozycja wobec zasad językoznawstwa strukturalnego. Wyrażany jest postulat wzbogacenia teorii semantyki języka naturalnego przez badanie sposobów pojęciowej (konceptualnej) organizacji wiedzy o świecie w wypowiedziach i znaczeniach jednostek językowych. Chodzi o opracowanie takich modeli przetwarzania informacji językowej przez człowieka, które stanowiłyby łącznik między teorią poznania, psychologią i teoriami semantycznymi. Zakłada się, że w mózgu istnieją tzw. konceptualne struktury informacji, w ramach których informacja pochodząca z komunikacji językowej staje się kompatybilna z informacją pochodzącą z systemów wzroku, słuchu, węchu, kinestezji, dzięki czemu człowiek potrafi komunikować o swoich doznaniach.

Językoznawstwo kognitywne „powstało w latach siedemdziesiątych w USA, od samego początku zyskując sobie wśród językoznawców zarówno entuzjastycznych zwolenników jak i nieprzejednanych krytyków. Idee, które jedni uważali za fascynujące, inni zaś za szalone i pozbawione naukowych podstaw, szybko przedostały się przez ocean, docierając do europejskich ośrodków językoznawczych. (...) Językoznawcom-kognitywistom przyświeca oczywiście taki sam cel, jaki stawiają sobie lingwiści innych orientacji: ma to być adekwatny opis zjawisk językowych, prowadzący do ich wyjaśnienia i ostatecznie pozwalający na ich trafne przewidywanie. Klucza do realizacji tego programu badawczego szukają w tym, co uznają za źródło, a zarazem rację istnienia języka: w zdolnościach ludzkiego umysłu, dzięki którym obserwacja otaczającego świata ostatecznie prowadzi człowieka do zdobywania wiedzy o tym świecie. Kluczem do zrozumienia natury ludzkiego języka jest zatem zrozumienie natury procesów ludzkiego poznania”<sup>56</sup>.

---

<sup>54</sup> Lindsay 1984, s. 520.

<sup>55</sup> Inne aspekty tej problematyki będą omówione w rozdz. drugim.

<sup>56</sup> Tabakowska 1995, s. 3, 5.

Za umowny początek kognitywizmu w nauce o informacji przyjmuje się rok 1977, kiedy to w Gandawie odbyły się międzynarodowe warsztaty International Workshop on the Cognitive Viewpoint. Mianem kognitywnych objęto wówczas badania skupione wokół „konceptualnych systemów użytkowników informacji”<sup>57</sup>. Obecnie za kognitywne uznawane są tak różne typy badań i aktywności poznawczej, że można przypuszczać, iż pierwotne znaczenie tego terminu w nauce o informacji zostało w znacznym stopniu zmienione.

Na warsztatach tych Belkin określił kierunek działania kognitywistów jako dążenie do doskonalenia przepływów informacji, szczególnie poprzez tworzenie modeli wiedzy kompatybilnych z ramami pojęciowymi używanymi w systemach informacyjno-wyszukiwawczych.

Peter Ingwersen<sup>58</sup> widział dwa niezależne strumienie badań, które po połączeniu mogłyby utworzyć, jak to nazwał, paradygmat kognitywny w projektowaniu systemów informacyjno-wyszukiwawczych. Pierwszym strumieniem jest tradycyjne (czyli takie jak do tej pory) rozwijanie i doskonalenie siw. Drugi strumień to, ogólnie mówiąc, badania użytkowników – potrzeb i problemów, zachowań informacyjnych oraz interakcji między użytkownikami a pośrednikami. Synteza tych dwóch strumieni powinna utworzyć paradygmat kognitywny, w którym, zdaniem Ingwersena, projektowanie systemu informacyjno-wyszukiwawczego koncentrowałoby się na interakcyjnym systemie z bazą wiedzy, uwzględniającym w swoim działaniu zarówno funkcje samego systemu, jak i funkcje tych, którzy z niego korzystają.

Prób określenia kierunków aktywności kognitywnej w informacji podejmowali się różni autorzy; przegląd literatury na ten temat można znaleźć między innymi w pracach Bryce’a Allena<sup>59</sup>, Rudy Ramsey i Jacka Grimesa<sup>60</sup>, Christine Borgman<sup>61</sup>, Brendy Dervin i Michaela Nilana<sup>62</sup> oraz Lindy Smith<sup>63</sup>.

Badania kognitywne w nauce o informacji korzystają z metodologii i ram<sup>64</sup> objaśniających utworzonych w naukach kognitywnych, zwłaszcza w psychologii kognitywnej i sztucznej inteligencji. Procesy kognitywne, takie jak:

<sup>57</sup> Allen 1991a.

<sup>58</sup> Ingwersen 1987.

<sup>59</sup> Allen 1991a.

<sup>60</sup> Ramsey 1983.

<sup>61</sup> Borgman 1986a, b, 1984.

<sup>62</sup> Dervin 1986.

<sup>63</sup> Smith 1987.

<sup>64</sup> Aby wyjaśnić znaczenie wyrazu *rama*, który w uwzględnionej literaturze przedmiotu używany jest z dość dużą dowolnością, pokażemy dwa sposoby rozumienia tego terminu – w sztucznej inteligencji i w językoznawstwie. Opis pierwszego został utworzony na podstawie Artowicz 1990, zaś opis drugiego – na podstawie: Tokarski 1996. (1) Teoria ram i skryptów utworzona przez Minsky’ego, rozwijana między innymi przez Schanka i Abelsona, jako podstawowy schemat reprezentacji wiedzy sytuacyjnej została przedstawiona w dwu wersjach. Najpierw *rama*

uczenie się, zapamiętywanie, rozumienie, formułowanie i rozwiązywanie problemów, podejmowanie decyzji itp. oraz obiekty mentalne (pojęcia, idee, wiedza) będące przedmiotami tych procesów pozostają od blisko dwudziestu lat w kręgu zainteresowania kognitywistów. Badania empiryczne sposobów rozwiązywania problemów i formułowania pytań stały się podstawą stworzenia teorii przetwarzania informacji. Na podstawie wyników badań sformułowano też wnioski dotyczące między innymi reprezentacji wiedzy, efektywności różnych strategii rozwiązywania problemów, roli i istoty ekspertyzy w warunkach wykonywania złożonych zadań.

W nauce o informacji badania kognitywne prowadzone są niemal we wszystkich jej obszarach, choć głównym polem odniesień pozostają użytkownicy. Czasami rażą naiwnym mentalizmem<sup>65</sup>. Także i tu istotę kognitywnego podejścia określa koncepcja percepcji ludzkiej, poznania oraz struktur informacji czy wiedzy. Celem kognitywnego podejścia do zachowań związanych z poszukiwaniem i wykorzystaniem informacji jest ujawnienie, w jaki sposób obrazy, wizje i ramy<sup>66</sup> ludzi reprezentujących różne sfery działalności, pracy, zadań itp. wiążą się z dostępnością informacji, w jaki sposób wybór informacji jest zdeterminowany (określony) przez obraz (wizję, ramę), a także jak informacja może zmieniać ten obraz (wizję), lub inaczej – wpływać na ramę referencyjną.

Ogólnie mówiąc, kognitywne badania użytkowników koncentrują się na problemach: powstawania potrzeb informacyjnych, przetwarzania pytań przez

---

oznaczała pewną ogólną strukturę, umożliwiającą odwzorowanie dystynktywnych cech obiektów pozajęzykowych oraz sytuacji, których elementami są te obiekty. Obiekt reprezentowany był przez nazwę ramy, zaś jego atrybuty przez tzw. klatki (ang. *slots*) powiązane ze znaczeniem nazwy ramy relacjami paradygmatycznymi i/lub skojarzeniowymi. Nazwom ram (obiektów) oraz ich atrybutów (klatkom), interpretowanym jako zmienne przybierające określone wartości w charakterystykach obiektów nazywanych deskrypcjami, przypisane były reguły wnioskowania, a więc elementy wiedzy proceduralnej. Dzięki temu cechy dystynktywne pewnych obiektów mogły być automatycznie przypisywane (przez przeniesienie) opisom innych obiektów, należących do tej samej klasy. Zmodyfikowana wersja teorii ram i skryptów opierała się na założeniu, że skrypty i ramy są elementami struktur pamięci dynamicznej (w rozumieniu Schanka), posiadającej różne poziomy i reorganizującej się pod wpływem przetwarzania informacji: przypominania, zapamiętywania i generowania nowej informacji. (2) W językoznawstwie rozmaite nurty gramatyki kognitywnej zakładają w analizach semantycznych istnienie abstrakcyjnych wzorców (modeli) pojęciowych ewokowanych przez jednostki leksykalne: interpretacyjnych ram pojęciowych (Minsky, Fillmore, Taylor), idealizowanych modeli kognitywnych (Lakoff), różnego rodzaju domen (Langacker, Fife). Rama interpretacyjna będzie zatem kulturowo zinterpretowanym (sprofilowanym) obrazem świata (jego fragmentu) skorelowanym z daną jednostką leksykalną. Składają się nań obok bardzo stabilnego rdzenia semantycznego także cechy konotacyjne, zarówno te o dużym stopniu utrwalenia, jak również konotacje słabe, ale potencjalnie przewidywane przez wewnętrzną logikę transformacji cech semantycznych.

<sup>65</sup> Krytykę mentalizmu w podejściu do języków informacyjno-wyszukiwawczych można znaleźć między innymi w publikacjach: Blair 1990; Frohmann 1990.

<sup>66</sup> Terminy *obraz*, *wizja świata*, *rama* (referencyjna, odniesienia) niosą ze sobą założenie, że sposób rozumienia pojęć przez osobę oddziałuje na jej zachowanie, gdyż zalczy ono od tego, jak pojęcia te są tworzone, kształtowane i zmieniane.



ludzi, wzorach negocjacji, interfejsie w układach człowiek-system i wykorzystywaniu informacji. Sytuacje informacyjno-wyszukiwawcze są traktowane jako zadania rozwiązywania problemów, wysoko interaktywne, z ludźmi (i „uczłowieczonymi” komputerami) jako aktywnymi podmiotami. Od czasu pracy Roberta Taylora<sup>67</sup>, nie zważając na wnioski płynące z codziennej pracy bibliotekarzy, pytanie użytkownika jest traktowane jako wyraz aktualnej potrzeby informacyjnej i zarazem podstawa realizowania zasady najlepszego dopasowania<sup>68</sup>.

Najważniejsze elementy kognitywne związane z negocjacjami w układach „użytkownik systemu-pośrednik” i „szukający-interfejs” dotyczą spostrzegania (percepcji), wywoływania z pamięci i rozpoznawania pojęć, przetwarzania informacji, uczenia się, rozwiązywania problemów, podejmowania decyzji, a z punktu widzenia pośrednika dodatkowo komunikowania się i zachowań interpersonalnych. Skoro jednak systemy informacyjno-wyszukiwawcze nie są w stanie w pożądanym stopniu dostosować się do indywidualnych i zmiennych oczekiwań, to umysł ludzki musi dostosować się do systemu, biorąc za punkt wyjścia zrozumienie, do czego ma służyć wyszukana informacja. Stąd wywodzi się, na przykład, koncepcja taktyk wyszukiwawczych opracowana przez Bates<sup>69</sup>.

Z ogromu literatury dotyczącej kognitywizmu w informacji wybraliśmy te wątki, które wiążą się z problematyką języka informacyjno-wyszukiwawczego i jego używania. Za Allenem<sup>70</sup> przedstawimy je w podziale na badania dotyczące wiedzy o świecie, temacie, zadaniu i samym systemie informacyjno-wyszukiwawczym. Trzeba jednak zauważyć, że wyodrębnione cztery kategorie wiedzy mają w jakimś stopniu charakter arbitralny, gdyż w praktyce systemów informacyjno-wyszukiwawczych nie da się nakreślić ostrych granic między nimi.

## WIEDZA O ŚWIECIE

Z literatury wynika, że obszar ten nie był dotąd kompleksowo zbadany. Ograniczona jest zarówno liczba badań dotyczących tego, jak wiedza o świecie, zwana też wiedzą ogólną, może wpływać na wyszukiwanie infor-

---

<sup>67</sup> Taylor 1968.

<sup>68</sup> Wersig i Hennings przeprowadzili badanie eksperymentalne oparte częściowo na koncepcji powstawania i rozwiązywania problemów, a częściowo na skomputeryzowanych konceptualnych mapach kwereńdy i tekstu dokumentu. Ich celem było uniknięcie wad zasady najlepszego dopasowania poprzez porównywanie map powstających w wyniku probabilistycznych metod poszukiwania, tj. nie boole'owskich. O potrzebach informacyjnych, zaletach i wadach zasady najlepszego dopasowania można przeczytać między innymi w publikacjach: Belkin 1982; Ingwersen 1982; Próchnicka 1991; Wilson 1981; Woźniak 1989.

<sup>69</sup> Bates 1979a, b.

<sup>70</sup> Allen 1991a.

macji, jak i badań mających na celu wskazanie, jak wiedza o świecie może być wyodrębniana i efektywnie wykorzystana w siw. Przyпуска się, że wiedza o świecie w pewien sposób oddziałuje na organizację trzech pozostałych rodzajów wiedzy<sup>71</sup>. Z analizy tego, co w literaturze przedmiotu uznawane jest za wiedzę o świecie, wynika, że do tej kategorii wiedzy włączana jest między innymi, a może przede wszystkim, gdyż w kontekście tym niemal w ogóle nie spotyka się cytowań bibliograficznych badań wiedzy o świecie rozumianej jako rzeczywistość odwzorowywana w dokumentach źródłowych systemu, wiedza o użytkowniku, w tym o jego cechach demograficznych i umiejętnościach poznawczych. Kategoria wiedzy o świecie obejmuje zatem ogólną wiedzę poznawczą użytkownika systemu determinowaną przez różnorodne czynniki społeczno-psychologiczne oraz wiedzę będącą podstawą umiejętności logicznego rozumowania, w tym reguły wnioskowania.

W literaturze dotyczącej użytkowników od ponad ćwierćwiecza daje się wyraz przekonaniu, że zachowania ludzi związane z poszukiwaniem informacji (zachowania informacyjne) podlegają wpływom czynników wynikających z przynależności do określonych systemów społecznych: kulturowego, prawnego, ekonomicznego, społecznego, zawodowego, edukacyjnego itp. Czynniki te kształtują potencjał poznawczy człowieka, jego postawy, sposoby interpretowania rzeczywistości, wyznaczania celów, postrzegania i definiowania problemów, ich rozwiązywania itp. Słuszność tych poglądów potwierdza także współczesna psychologia poznawcza. Do badań wiedzy o świecie w nauce o informacji zalicza się na przykład analizy zależności pomiędzy poziomem wykształcenia i indywidualnym doświadczeniem w korzystaniu z techniki informacyjnej (ang. *information technology*) a rozumieniem świata oraz gotowością i umiejętnością zaspokajania potrzeb informacyjnych<sup>72</sup>, pomiędzy cechami etnicznymi czy płcią a stosunkiem do komputerów<sup>73</sup>. Wydaje się, że cechy o charakterze demograficznym na pewno nie mogą być uważane za silne przesłanki do wyrokowania o zachowaniach informacyjnych, choć niewątpliwie mogą i powinny być uwzględniane w budowaniu modeli użytkowników. Istotną sugestią projektową wynikającą z badań wiedzy o świecie jest postulat projektowania opcjonalnych mechanizmów interakcji system-użytkownik (tzw. interfejsów), zakładających różne poziomy wiedzy i umiejętności użytkowników w zakresie posługiwania się techniką informacyjną. Zagadnienie to wiąże się z problematyką organizacji i prezentacji wiedzy o systemie informacyjno-wyszukiwawczym.

---

<sup>71</sup> Allen 1991a, b.

<sup>72</sup> Barnett 1988.

<sup>73</sup> Allwood 1990; Krcndl 1989; Parasuraman 1990; Templc 1989.

## WIEDZA O SYSTEMIE INFORMACYJNO- -WYSZUKIWAWCZYM

Przy podejściu tradycyjnym użytkownik musi dysponować określoną wiedzą o systemie informacyjno-wyszukiwawczym, żeby móc efektywnie wyszukiwać w zbiorach siw. Często wiedza ta decyduje o powodzeniu całego wyszukiwania<sup>74</sup>, choć istnieją i takie badania, z których wynika, że różna wiedza wyjściowa użytkownika o siw nie ma, lub może nie mieć, wpływu na efekty wyszukiwania<sup>75</sup>. Na innym założeniu opiera się podejście kognitywne, w którym przyjmuje się, że to system powinien się dostosowywać do użytkownika, z jednej strony maksymalizując możliwości pozyskiwania wiedzy o siw bezpośrednio z samego systemu, z drugiej zaś minimalizując jej zakres.

Z analizy tego, co w literaturze przedmiotu uznawane jest za wiedzę o systemie, wynika, że ta kategoria wiedzy obejmuje (lub powinna obejmować): wiedzę będącą gwarancją umiejętności współpracy z systemem informacyjno-wyszukiwawczym, zwłaszcza skomputeryzowanym; wiedzę o zasobach informacyjno-wyszukiwawczych systemu oraz wiedzę o sposobach kodowania i organizowania w siw wiedzy o temacie i wiedzy o zadaniu.

Najbardziej owocnymi badaniami z zakresu wiedzy o siw były analizy pytań użytkowników i ich systemowych reprezentacji oraz sekwencji komend. Badania te zostały wykorzystane między innymi do wykrywania błędów i następnie oferowania pomocy dostosowanej do sytuacji, w jakiej znalazł się użytkownik<sup>76</sup>. Analizy instrukcji sformułowanych przez użytkowników pozwoliły ustalić typowe błędy formalne oraz zaprojektować mechanizmy ich usuwania współpracujące z użytkownikiem przy weryfikacji formuły wyszukiwawczej zanim zostanie wykorzystana w wyszukiwaniu.

Postulat wyposażania systemu informacyjno-wyszukiwawczego w wiedzę o nim samym jest dziś realizowany najczęściej w postaci różnych dostępnych online systemów pomocy (tzw. helpów) oraz tzw. inteligentnych interfejsów komunikacji system-użytkownik. W obu przypadkach najczęściej chodzi o niwelowanie trudności technicznych procesu wyszukiwania informacji. Znajduje to wyraz na przykład w zastępowaniu komunikacji za pomocą komend rozbudowanymi menu dopuszczalnych działań użytkownika. Coraz powszechniej stosuje się też różnego rodzaju mnemotechniczne metody prezentowania informacji o funkcjach systemu, na przykład w postaci piktogramów lub wyrażen języka naturalnego łatwo kojarzących działania systemu z sytuacjami życia codziennego.

---

<sup>74</sup> Borgman 1986b; Katzeff 1990.

<sup>75</sup> Liebscher 1988.

<sup>76</sup> Canter 1985.

Zbyt często jednak wiedza o systemie sprowadzana jest jedynie do wiedzy o aspektach technicznych funkcjonowania siw. Badania kognitywne zdają się charakteryzować tendencją do ignorowania takich elementów systemów informacyjno-wyszukiwawczych jak zakres i rodzaje dokumentów źródłowych systemu, język informacyjno-wyszukiwawczy i zasady jego stosowania, organizacja zbioru informacyjno-wyszukiwawczego itp.

W literaturze rozpowszechniona jest opinia, że istnieje ścisły związek między środkami technicznymi systemu a językiem informacyjno-wyszukiwawczym. Tymczasem praktyka pokazuje, że jak do tej pory środki techniczne zdają się mieć wpływ tylko na formy prezentacji oraz sposoby realizacji języków informacyjno-wyszukiwawczych, nie determinując znacząco ich struktury wewnątrzjęzykowej. Nie jest naszym zamiarem pomniejszanie wagi środków technicznych, ale nie możemy zgodzić się z poglądem o ich wyjątkowym znaczeniu i wpływie na efekty wyszukiwania. Środki techniczne mogą mieć dodatkowy pozytywny wpływ na efektywność wyszukiwawczą, jeśli inne elementy systemu, zwłaszcza te o charakterze intelektualnym, mają odpowiednio wysoką jakość. Należy przy tym pamiętać, że: (1) język informacyjno-wyszukiwawczy systemu nie istnieje w izolacji; przeciwnie, powinien stanowić „naczynie połączone” z repertuarem technik i środków wyszukiwawczych, jakim dysponuje dany system (wagi, maskowanie, operatory Boole’a, operatory kontekstowe itd.); (2) w zautomatyzowanych systemach informacyjno-wyszukiwawczych coraz częściej rezygnuje się z ograniczania wyszukiwania rzeczowego tylko do elementów rekordu specjalnie do tego celu przeznaczonych (pole haseł przedmiotowych, symboli klasyfikacyjnych czy deskryptorów), powszechnie korzystając z możliwości przeszukiwania wszystkich lub przynajmniej wielu pól rekordu; na zjawisko to dodatkowo nakłada się tendencja do wzbogacania opisów o takie, nietypowe dla systemów tradycyjnych, elementy wyszukiwawcze jak na przykład adnotacje, analizy dokumentacyjne, terminy pobierane ze spisów treści czy indeksów przedmiotowych katalogowanych dokumentów, czasami całe spisy i indeksy<sup>77</sup>; (3) odpowiednie oprogramowanie stwarza możliwości zapoznawania się z przebiegiem wyszukiwań zrealizowanych przez użytkowników, z jego etapami, doбором terminów, strukturą wyrażenń złożonych itd., co z kolei może i powinno być wykorzystane do optymalizowania mechanizmów siw; (4) istotnie zmienia się miejsce słownika jiw w katalogu online. W systemach tradycyjnych słownik języka informacyjno-wyszukiwawczego, najczęściej w formie drukowanej lub kartotekowej, istnieje, nazywając obrazowo, poza bazą, w której jest wykorzystywany. W wielu nowoczesnych systemach informacyjno-wyszukiwawczych automatyzacja nie wiąże się z integracją słownika, który nadal pozostaje w swej tradycyjnej postaci fizycznej poza bazą danych lub, w najlepszym razie,

---

<sup>77</sup> Obserwując współczesne systemy informacyjno-wyszukiwawcze można dostrzec tendencję do zmierzania w kierunku systemów pełnotekstowych (zwanych obecnie także archiwami elektronicznymi), chociaż owa pełnotekstowość różnie bywa rozumiana.

zostaje zapisany w pamięci systemu jako wyizolowany jego element, do którego sięganie wymaga dodatkowej wiedzy i aktywności. Jest to chyba jeden z najbardziej drastycznych przykładów niewykorzystania w systemie możliwości stwarzanych przez technikę i postulatów będących dorobkiem poznania kognitywnego. Integracja słownika jiw z obsługiwaną bazą danych wniosłaby szereg usprawnień do procesu indeksowania i wyszukiwania, nadała nową jakość temu narzędziu wyszukiwawczemu, zmieniając jego funkcję i częściowo zawartość. Zawartość słownika zintegrowanego można podglądać na każdym etapie procesu wyszukiwawczego, modyfikując, w miarę uzyskiwania kolejnych informacji, wyjściową instrukcję wyszukiwawczą. Informacje ze słownika jiw wzbogacają indeks zbioru wyszukiwawczego, a zbiór wyszukiwawczy wzbogaca słownik jiw na przykład o liczniki wystąpień wyrażeń jiw w bazie danych mówiące, ile dokumentów jest zaindeksowanych danym terminem jiw, co umożliwia m.in. przyjęcie odpowiedniej strategii wyszukiwawczej.

## WIEDZA O ZADANIU

Zachowaniem użytkowników w procesach szukania informacji kierują też cele, do realizacji których informacja jest potrzebna. Wiedza o zadaniu, zwana też wiedzą o problemie wyszukiwawczym lub problemie informacyjnym, może mieć istotny wpływ na efekty wyszukiwania. Tradycyjnie wiedza o zadaniu łączona jest tylko z użytkownikami, podczas gdy porównywanie wiedzy użytkownika i wiedzy odwzorowanej w systemie wymaga jednak wyodrębnienia analogicznej struktury w systemie informacyjno-wyszukiwawczym. Wiedzą systemu dokumentacyjnego o problemie użytkownika jest zbiór metainformacji, które system uznaje za relewantne dla użytkownika na podstawie interpretacji systemowej reprezentacji wiedzy o dziedzinie z punktu widzenia pytania (pytań) informacyjnego reprezentowanego w systemie przez jego instrukcję wyszukiwawczą.

Znaczącymi badaniami wiedzy o zadaniu (różnie nazywanej przez różnych autorów) są prace między innymi Taylora<sup>78</sup>, Dervin<sup>79</sup> i Belkina<sup>80</sup> omawiane w literaturze polskiej przez Jadwigę Woźniak<sup>81</sup> i Marię Próchnicką<sup>82</sup>. Na uwagę zasługują zwłaszcza badania, które zidentyfikowały i usystematyzowały cele użytkowników i ich plany. Najwięcej badań z obszaru wiedzy o zadaniu dotyczyło analizy opisu sytuacji problemowej użytkownika. Wśród nich do najbardziej znanych należą badania Belkina<sup>83</sup> dotyczące „niewystarczającego

---

<sup>78</sup> Taylor 1986, 1985, 1984.

<sup>79</sup> Dervin 1986, 1983, 1980, 1977.

<sup>80</sup> Belkin 1990, 1986, 1984, 1983, 1982.

<sup>81</sup> Woźniak 1989.

<sup>82</sup> Próchnicka 1991.

<sup>83</sup> Belkin 1987.

stanu wiedzy”, będące częścią szerszej zaplanowanych badań z zakresu modularnego podejścia do projektowania systemów ekspertowych<sup>84</sup>. Inne podejście do modelowania problemu użytkownika prezentuje system THOMAS<sup>85</sup>, w którym cele i intencje użytkownika są modelowane w postaci starannie wyselekcjonowanych i ważonych segmentów modelu świata, odwzorowującego związki między dokumentami, autorami i terminami. Siw operujące modelami wiedzy o zadaniu, podobnie jak w wypadku wielu innych modeli kognitywnych, miały na ogół charakter eksperymentalny.

Celem wszystkich tych badań, podejść i koncepcji jest wypracowanie metody identyfikacji wiedzy o zadaniu na podstawie tego, co użytkownik postrzega jako podstawowe wymiary sytuacji problemowej. Według Taylora problem wyszukiwawczy determinowany jest przez oceny wartości różnych własności informacji w różnych sytuacjach problemowych. Według Dervin identyfikacja luki informacyjnej (ang. *gap-use, situation-gap-use*) wiąże się ze stwierdzeniem niespójności indywidualnej semantycznej interpretacji wyrażen, za pomocą których opisuje się pewien fragment rzeczywistości i konceptualnego odwzorowania rzeczywistości w prywatnej wiedzy użytkownika. Reprezentację wiedzy o zadaniu ma stanowić opis sytuacji ujawnienia się tej niespójności. Z kolei zdaniem Belkina identyfikacja problemu wyszukiwawczego powinna mieć charakter dynamicznego procesu negocjacyjnego między użytkownikiem i systemem, w którym określone zostaną między innymi cele użytkownika, faza rozwiązywania problemu, stopień jego strukturalizacji, stan wiedzy o temacie, konceptualne odwzorowanie problemu, poziom i profil wykształcenia użytkownika oraz jego doświadczenie w korzystaniu z systemu.

Wiedza o zadaniu charakteryzuje się niepewnością i niepełnością. Oddziałują nań zarówno, niekiedy bardzo małe, struktury wiedzy użytkownika, jak i elementy o charakterze sytuacyjnym (kontekstowym), na przykład cechy osobnicze użytkownika, cele, motywacje podejmowanych zachowań informacyjnych. Stąd naturalny wniosek, że wiedza o zadaniu ma charakter dynamiczny, podlegając nieustannym zmianom w wyniku interakcji użytkownika z systemem informacyjno-wyszukiwawczym.

## WIEDZA O TEMACIE

Na wyniki wyszukiwania istotny wpływ (prawdopodobnie większy niż wiedza o świecie i wiedza o systemie) ma wiedza użytkownika o temacie poszukiwania lub zakresie przedmiotowym, w którym temat ten się lokuje. Jest ona określana mianem wiedzy o temacie, domenie lub dziedzinie. Najczęściej badana jest w dwóch aspektach – poziomu wiedzy i jej organizacji<sup>86</sup>. Ale

<sup>84</sup> Belkin 1986, 1983; Brooks 1986.

<sup>85</sup> Oddy 1977; Ofori-Dwumfuo 1984.

<sup>86</sup> Allen (Allen 1991b) wykazał, że poziom wiedzy użytkownika o temacie wpływa na sposób, w jaki formułuje on instrukcje wyszukiwawcze, przeszukując katalog online.

wiedza o temacie to nie tylko wiedza użytkownika; jest to także zapisana w zbiorach informacyjno-wyszukiwawczych systemu wiedza o rzeczywistości odwzorowanej w dokumentach źródłowych systemu.

Sposoby modelowania tej kategorii wiedzy zależą od wielu czynników, z których najważniejszymi są własności systemu informacyjno-wyszukiwawczego (m.in. zautomatyzowany czy manualny) i metainformacyjnej reprezentacji dokumentów w systemie. Jednym ze sposobów modelowania wiedzy o temacie są asocjacyjne sieci semantyczne<sup>87</sup>. Przy tej technice wiedza ekspercka o domenie jest reprezentowana za pomocą słownika, w którym poszczególne jednostki leksykalne pozostają wobec siebie w bliskości lub oddaleniu stosownie do „odległości semantycznej”, jaka między nimi występuje. Taka sieć może być uaktywniana w celu ustalenia słownictwa najbliższego wyrażeniom wyszukiwawczym wprowadzonym przez szukającego<sup>88</sup>.

Jak różne są poziomy wiedzy o temacie właściwe różnym użytkownikom, tak różna też jest jej organizacja, mająca niewątpliwie wpływ na wyszukiwanie informacji. Zdarza się, że wiedza systemu o temacie odwzorowana w postaci sieci semantycznej czy tezaury zorganizowana jest inaczej niż wiedza użytkownika o tej samej domenie. Możliwym rozwiązaniem tego problemu mogłoby być kilkakrotne indeksowanie za pomocą odpowiedniej liczby różnych języków informacyjno-wyszukiwawczych o sieciach relacji wyekstrahowanych z modeli eksperckich i modeli wiedzy użytkowników o temacie, inaczej mówiąc, „sugerowanych” przez różne grupy użytkowników<sup>89</sup>. Stosunkowo często stosowanym sposobem wykorzystania wiedzy o temacie jest włączanie do interfejsu mechanizmu wskazującego przykłady (ang. *query by example*), które mogą być pozyskiwane w różny sposób, np. mogą być wskazane przez specjalistę, generowane przez specjalne programy symulacyjne, uznane przez samego użytkownika za relewantne itp.

Wiedza o temacie jest oczywiście bardzo ważna dla wyszukiwania informacji. Badania projektowe w tym obszarze koncentrowały się dotychczas na modelowaniu wiedzy w sposób optymalny z punktu widzenia użytkowników

---

<sup>87</sup> Mianem *sieci semantyczne* określa się obecnie różne sposoby reprezentacji wiedzy, których jedyną wspólną cechą jest podobieństwo strukturalne notacji. Bazę notacji stanowią węzły, graficznie oznaczane jako kropki, kółka lub prostokąty oraz łuki, graficznie oznaczane strzałkami. Zarówno węzły, jak i łuki są opatrywane etykietami, pełniącymi funkcję wykładników znaczenia wyrażenń składowych sieci. Etykiety węzłów reprezentują wyrażenia elementarne języka desygnujące obiekty, stany, zdarzenia itp., natomiast łuki odwzorowują relację między faktami desygnowanymi przez wyrażenia elementarne. W literaturze przedmiotu podkreśla się, że poza podobieństwem notacji trudno jest wskazać uniwersalne zasady interpretacji sieci semantycznej. Ich charakterystyka jako sposobu reprezentacji wiedzy ogranicza się bądź do prezentacji rozwiązań już historycznych, bądź do najczęściej stosowanych zasad organizacji pola semantycznego języka. W nowszych pracach na temat sieci semantycznych obserwuje się tendencję do łączenia sieci w złożone struktury nazywane ramami (ang. *frames*), scenariuszami (ang. *scenarios*) czy skryptami (ang. *scripts*), w których wykorzystuje się także proceduralne metody reprezentacji wiedzy.

<sup>88</sup> Croft 1986.

<sup>89</sup> Pejtersen 1987.

potrzebujących danej informacji. Zwykły tezaurus może służyć jako prosty model tej wiedzy. Bardziej wyrafinowane zadania mogą realizować na przykład sieci semantyczne. Mają one zdolność kierowania użytkownika do precyzyjnie określonych obszarów tematycznych, sugerowania dodatkowych tematów oraz lokowania poszczególnych tematów wyszukiwawczych indywidualnych użytkowników w obrębie większych domen tematycznych.

Organizacja i prezentacja semantycznych kategorii leksykalnych należy do najbardziej typowych zastosowań sieci semantycznych w dokumentacyjnych systemach informacyjno-wyszukiwawczych. Znaczenie i forma jednostek leksykalnych reprezentowane są przez węzły sieci, łuki między węzłami pojęciowymi reprezentują relacje paradygmatyczne wiążące określone jednostki, zaś łuki między węzłami pojęciowymi i węzłami wyrażeniowymi reprezentują związek wyrażania (określone znaczenie jest wyrażane przez określone formy językowe). Ten sposób wykorzystania sieci semantycznej ilustruje na przykład system MenUSE<sup>90</sup>. Sieć semantyczna MenUSE stanowi reprezentację wiedzy o słownictwie zgromadzonym w Medical Subject Headings i stosowanym w bazach MEDLINE. Działanie systemu MenUSE (i innych podobnych) kolejny raz potwierdza wcześniej sformułowaną obserwację, że rozwój techniki informacyjnej ma proporcjonalny wpływ na efektywność wyszukiwania informacji i satysfakcjonowanie użytkownika tylko wtedy, gdy sposoby i środki reprezentacji (język informacyjno-wyszukiwawczy, jego struktura semantyczna, w tym struktura semantycznych kategorii leksykalnych, pragmatyka jego stosowania, jakość pracy indeksatorów itp.) i wyszukiwania informacji charakteryzują się odpowiednią jakością i relewantnością.

## OSOBISTE SYSTEMY INFORMACJI JAKO PRZYKŁAD PODEJŚCIA KOGNITYWNEGO

Ekspansja komputeryzacji, w tym zdalnego dostępu do zasobów informacji, spowodowała istotną zmianę w zachowaniach informacyjnych użytkowników. Coraz większa ich liczba samodzielnie poszukuje informacji w licznych bazach danych. Oczywiście nadal istnieją grupy użytkowników korzystających z pomocy pośredników, na przykład bibliotekarzy. Są to ci, którzy albo nie potrafią, albo nie muszą sami realizować wyszukiwań. W literaturze i w praktyce uwagę koncentruje się jednak na tzw. użytkownikach ostatecznych (końcowych, ang. *end users*). Z myślą o nich projektuje się narzędzia usprawniające wyszukiwanie informacji, dbając by były „przyjazne”. Mieści się w tym także „przyjazność” systemów kategorii w jiw, której jednym z wyznaczników jest budowanie systemu kategorii wokół osi kategorii najbardziej użytecznych (podstawowych) dla określonej grupy użytkowników. W niektórych systemach

---

<sup>90</sup> Pollitt 1990.



języki informacyjno-wyszukiwawcze są w większym stopniu podporządkowane wymogowi przyjazności wobec użytkownika niż zasadzie dobrego reprezentowania dokumentów.

Systemy kategoryzacji i klasyfikacji służą organizowaniu informacji, nakładając na informację pewną strukturę (w przypadku kategoryzacji empirycznej), explicite wyznaczającą relacje między jej elementami (klasami). Nałożenie takiej struktury presuponuje, że wiedza posiada znane granice, że to, co zawiera się w tych granicach, można podzielić na odrębne, niezależne klasy i że każdy przedmiot ma swoje przyporządkowanie do klasy. Kategoryzacja realizuje zatem funkcje analogiczne do klasyfikacji piśmiennictwa, tj. definiuje (przez ustalanie klas jednostek dzielących specyficzne cechy czy atrybuty) oraz porządkuje (przez tworzenia układu czy układów klas wyrażających relacje konceptualne pomiędzy elementami struktury), choć eksponowanie układu klas jest w zasadzie specyficzne dla klasyfikacji piśmiennictwa (w odróżnieniu od klasyfikacji w logice i kategoryzacji).

Tradycyjny schemat hierarchicznej klasyfikacji bibliotecznej dostarcza konceptualnej struktury czy też ramy, która odzwierciedla relacje znaczeniowe między klasami. Każde pojęcie ma wyznaczone wejście w ramach klasy nadrzędnej oraz przypisany jednolity symbol notacji jednoznacznie je identyfikujący. Sztywność właściwa tradycyjnym schematom klasyfikacyjnym jest przyczyną, że możliwość ich reakcji na rozwój wiedzy jest ograniczona, zaś pewna powtarzalność klas może sprzyjać rozproszeniu informacji. Uważa się także, że hierarchiczna struktura schematów klasyfikacji bibliotecznych zawodzi w konfrontacji z wiedzą indywidualnych użytkowników związaną z doświadczaniem rzeczywistości. Wielu autorów podkreśla też niezdolność struktur tradycyjnych do sprostania idiosynkratycznym reprezentacjom wiedzy, pisząc o „różnicach” charakteryzujących postrzeganie świata przez jednostki: jaką terminologię stosują do opisu problemu; jak strukturalizują problem i identyfikują jego części składowe; jaki jest poziom złożoności informacji wymaganej do zaspokojenia ich zapytania itp. W tym ujęciu fundamentalnym problemem stojącym przed nauką o informacji jest konieczność umożliwienia skutecznej interakcji między użytkownikiem i systemem, interakcji, która sprzytałaby dostępowi do odpowiednich i relewantnych materiałów w ramach ograniczeń nałożonych przez kognitywne wyobrażenia indywidualnego użytkownika. Zrozumienie tego związku musi się zacząć od samego procesu klasyfikacji i kategoryzacji, a zwłaszcza ich roli w poznaniu. Zbudowany ze zrozumieniem problemu użytkownika interfejs pośredniczący między językiem informacyjno-wyszukiwawczym a użytkownikiem może wyeliminować lub przynajmniej istotnie zredukować niezbędną wiedzę użytkownika o samym jiw. Zautomatyzowany system informacyjno-wyszukiwawczy może, i powinien, dysponować takimi sposobami i poziomami prezentacji języka, aby język ten mógł być efektywnie wykorzystywany przez różne kategorie użytkowników.

Zasady kategoryzacji w ujęciu psychologii kognitywnej mają szczególne zastosowanie w badaniu indywidualnych systemów informacji – prywatnych (osobistych) zbiorów dokumentów i informacji o nich. W środowisku akademickim zbiory prywatne mają istotny wpływ na zachowania informacyjne. Przeprowadzono badania<sup>91</sup>, z których wynika, że większość badaczy utrzymuje własne zbiory i traktuje je jako pierwotne źródło zaspokajania potrzeb informacyjnych. Aczkolwiek osobiste systemy porządkowania dokumentów i informacji o nich rzadko są przedmiotem analizy specjalistów od badań informacyjnych, faktycznie odgrywają ważną rolę w użytkowaniu informacji. Indywidualne przyzwyczajenia dotyczące obchodzenia się z dokumentami i informacją o nich mogłyby być źródłem cennych wskazówek dla projektantów systemów i języków informacyjno-wyszukiwawczych.

Analiza osobistych systemów informacji, zwanych też domowymi, rozpoczęła się od badania indeksów tworzonych i wykorzystywanych przez naukowców do odszukiwania informacji w zbiorach własnych. Interesujące jest spostrzeżenie, że tendencja do utrzymywania własnego systemu informacji nasila się wraz ze wzrostem pozycji zawodowej. Osobisty indeks można uznać za punkt wielostronnego dostępu do wyspecjalizowanego zbioru, zawierającego różnorodne materiały publikowane i niepublikowane, w tym nowo powstałe media, takie jak na przykład poczta elektroniczna czy czasopisma elektroniczne.

Uwagę przyciągają też sposoby kategoryzowania i korzystania ze zbiorów informacji i dokumentów przez ich właścicieli (naukowców). Znane są badania, w których analizowano fizyczną organizację takich zbiorów, ich zakres, treść, formy itd., a także cechy i własności, odróżniające je od zbiorów zinstytucjonalizowanych. Case określił poziomy gromadzenia (przestrzeń, forma, temat i przetwarzanie) i stwierdził, że przy organizowaniu dokumentów pierwszeństwo daje się przestrzeni fizycznej, następnie formie dokumentów<sup>92</sup>. Ustalenie to potwierdza wyniki wcześniejszych badań przeprowadzonych w 1976 roku przez Sopera, który analizował teorie dotyczące dostępności informacji. Dostępność fizyczna powodowała wzrost wykorzystania, stąd też osobiste zbiory były traktowane jako bardziej użyteczne.

Case wybrał do badania historyków, gdyż historycy w swojej pracy intensywnie wykorzystują dokumenty, żeby zaś wykonywać swą pracę efektywnie i sprawnie, muszą mieć biegle opanowane techniki rozpoznawania, przeglądania, indeksowania, archiwizowania i wyszukiwania dużych ilości tekstów wiążących się z ich przedsięwzięciami badawczymi. W porównaniu z innymi naukowcami praca historyka obejmuje szeroki zakres metod i tematów, co pozwala rozszerzyć wnioski dotyczące ich zachowań informacyjnych na

---

<sup>91</sup> Case 1991; Lundeen 1982; Soper 1976.

<sup>92</sup> Case 1991, s. 657-668.

zachowania badaczy kilku dziedzin wiedzy. Praca historyków pokrywa zaledwie małą część zagadnień badanych przez nauki społeczne, kładzie jednak nacisk na wymiar czasowy, który często jest ignorowany przez inne dyscypliny.

Mimo iż praca historyka jest w głównej mierze pracą poznawczą, to Case'a interesowały przede wszystkim wzajemne oddziaływania między mentalnym i fizycznym przetwarzaniem materiałów tekstowych – jak to się dzieje, że teksty wyselekcjonowane przez historyka jako „odnoszące się do badanego zagadnienia” są indeksowane i przechowywane w celu późniejszego odszukania. Inaczej mówiąc, w jaki sposób historyk zarządza – na drodze rozmieszczenia i etykietowania fizycznego oraz intelektualnego – dużymi ilościami tekstu, które gromadzi w procesie pracy.

Czynności wykonywane przez historyka z pewnością zmieniają się wraz z postępem prac nad tematem. Wszystkie stadia łączy jednak potrzeba zbierania (nieraz wyłącznie pamięciowego, lecz najczęściej fizycznego) informacji i przechowywania jej. Historycy badani przez Case'a przejawiali tendencję do przechowywania wielkich ilości zapisanych informacji. Podobnie jak w bibliotekach nadrzędnym celem badanych było przechowywanie podobnych przedmiotów razem, o ile pozwalały na to inne ograniczenia. Owe inne ograniczenia obejmowały takie wymogi jak: potrzeba trzymania niektórych rzeczy „pod ręką” (np. materiały często używane), konieczność zmieszczenia czegoś na ograniczonej przestrzeni, a także potrzeba zachowania funkcji przypominania o czymś (np. trzymanie materiałów na bieżące zajęcia w takim miejscu, gdzie nie sposób ich nie zauważyć).

Przechowywanie podobnych przedmiotów razem może oznaczać cały szereg rzeczy. Podobnie jak biblioteki, badacze również dążą do przechowywania dokumentów dotyczących zbliżonej tematyki tak blisko siebie, jak tylko się da. Jednak nad tą nadzwyczaj ważną regułą dominuje inna – często pożądane jest trzymanie podobnych form blisko siebie. Nie ulega wątpliwości, że niektóre formy fizyczne nie pasują do siebie. Trudno na przykład przemieszczać nadbitki artykułów z książkami, aczkolwiek dwóch respondentów robiło to, posługując się tekturowymi okładkami dla grup artykułów. Jak się zatem okazuje, przy tworzeniu porządku fizycznego forma w niektórych sytuacjach bierze górę nad treścią. Biblioteki także niekiedy przedkładają formę fizyczną nad treścią, grupując książki największe (ponadwymiarowe) i najmniejsze, oraz w sensie intelektualnym, tworząc reprezentacje niektórych dokumentów poprzez wysunięcie na plan pierwszy cech pozatreściowych.

Tak więc głównymi zasadami organizowania są czynniki fizyczne, względnie na ograniczenia przestrzenne, a zaraz potem forma zewnętrzna organizowanych obiektów. Kilku respondentów zadało sobie sporo trudu, by dokładnie zaindeksować swoje zbiory, ale najczęściej kartoteki tworzone przez historyków były bardzo proste i dotyczyły książek i artykułów z czasopism. Zgodnie z tym, co podali respondenci, wiele czynności organizowania infor-

macji zachodzi dopiero w trakcie pisania. Funkcją wielu kartotek i spisów skomputeryzowanych było uchwycenie pewnej sekwencji pomysłów, które miały być wykorzystane przy pisaniu określonej pracy.

Dopiero na trzecim poziomie organizowania pojawia się to, co normalnie uważamy za obiekt odniesienia intelektualnego – temat lub przedmiot. Respondenci zostali poproszeni o przespacerowanie się wokół swego gabinetu i opowiedzenie o tym, co tam się znajduje. Typowa wypowiedź wyszła z ust respondenta, którego cały księgozbiór znajdował się w jego gabinecie i był zorganizowany według tematyki: „(...) oraz w pewnym sensie chronologicznie... są tam rzeczy dotyczące historii Ameryki, a te kartoteki pośrodku dotyczą historii państwa i prawa... podręczniki, podręczniki ogólne historii Ameryki... następnie idę chronologicznie: okres kolonialny, koniec XIX w., wojna secesyjna, rekonstrukcja, czasy postępu, New Deal, zimna wojna, polityka zagraniczna... tak, układ jest nieco chronologiczny, a nieco chronologiczny i tematyczny”<sup>93</sup>. Inny respondent stosował następujące kategorie semantyczne: „cywilizacja zachodnia, literatura amerykańska, opracowania nt. literatury amerykańskiej, starodruki... duplikaty książek posiadanych w domu... opracowania ogólne nt. religii, socjologia, wiek dwudziesty, historia kobiet, historia kolorowych, podręczniki, różne, pisma Thomasa Jeffersona, historia Anglii”<sup>94</sup>.

Większość respondentów organizowała swe kartoteki głównie tematycznie. Pewna ich część uznała za potrzebne porządkowanie również według chronologii okresów lub wydarzeń, kombinacji tych dwóch kategorii lub tworząc niższe poziomy organizacji informacji i/lub dokumentów. Tylko dwóch historyków stwierdziło, że organizuje swe materiały wyłącznie chronologicznie, a jeden według tematu w obrębie okresu chronologicznego. Pozostali porządkowali kartoteki według nazwiska autora, co Case uznał za odmianę pośredniej organizacji tematycznej. Jeden z respondentów podzielił nawet posiadane materiały według kryteriów jakościowych („Dobre, niedawno napisane, prace z dziedziny historii kultury umysłowej Ameryki, złe, niedawno napisane, prace z dziedziny historii kultury umysłowej Ameryki”<sup>95</sup>). Wielu respondentów oddzielnie trzymało podręczniki, albo wewnątrz, albo poza szerszymi zgrupowaniami tematycznymi, dokonując podziału według przeznaczenia.

Podsumowując, praktycznie w każdym przypadku pojawiały się cztery poziomy gromadzenia:

- (przestrzeń (przestrzeń fizyczna, zależna od układu ścian, półek na książki, itd.);
- (forma (pojedyncza kartka, wiele kartek, książka w miękkiej oprawie, książka w twardej oprawie, inne media);

---

<sup>93</sup> Case 1991, s. 663.

<sup>94</sup> Ibidem s. 663.

<sup>95</sup> Ibidem s. 664.

- (temat (szeroko rozumiana tematyka, np. „sprawy zagraniczne” lub okres, np. „wczesny okres niepodległości”);
- (ujęcie (np. biografie), przeznaczenie (np. podręczniki, książki na zajęcia), jakość („dobre” przykłady w przeciwieństwie do „złych”).

Zauważmy analogię do studiów międzykulturowych poświęconych percepcji i nazywaniu kolorów, które wykazują, że czarny i biały to barwy, które jako pierwsze są powszechnie wyróżniane za pomocą nazwy, następnie czerwony, później zaś albo żółty, albo niebieski, albo zielony. Podobny trop daje się zaobserwować w pracowniach naukowców – priorytetowo przy rozplanowaniu lokalizacji dokumentów traktowana jest przestrzeń fizyczna, następnie forma fizyczna dokumentu. Dopiero trzeci poziom dotyczy tematu dokumentu, czynnika najbardziej interesującego badacza informacji.

Warte zauważenia jest, że prawie wszyscy badani<sup>96</sup> w swych wypowiedziach posługiwali się metaforami<sup>97</sup>, porównującymi idee z bytami fizycznymi. Na przykład:

- „pytania i pomysły pływają mi po głowie” (pomysły jako ryby)
- „rzucam swą sieć bardzo szeroko” (pomysły jako ryby)
- „siedzę i staram się to rozgryźć” (pomysły jako rebusy)
- „tnę i sklejam” (pomysły jako kawałki papieru)
- „staram się śledzić rzeczy” (kartotekowanie jako polowanie)
- „nakleić etykietę” (pomysły jako słoiki)
- „pożyczamy i kradniemy od siebie nawzajem bezustannie” (pomysły jako przedmioty cenne)
- „bawiłem się” pomysłami na kartach (pomysły jako zabawki)
- „zwalczyć literaturę” (pomysły jako przeciwnicy)
- „zadawać pytania materiałowi” (pomysły jako ludzie)
- „drażą temat” (uczeni jako krety)
- „moje zainteresowania są ograniczone czasowo” (tematy jako przestrzeń)
- „w głowie mam bibliografię bibliografii” (tematy jako spisy odnośnych dokumentów).

Zauważmy, że w większości są to, według terminologii Lakoffa i Johnsona, metafory idei – „pomysły jako organizmy” lub „pomysły jako produkty”. Cytowane metafory historyków amerykańskich należą do tych, które można by określić jako niezłożone. Inne wymagają dłuższego opisu. Jeden z respondentów na przykład opisał zmiany zaszłe w ostatnich czasach w historiozofii najpierw jako „przebudowę domu” (tj. rewizję tradycyjnych poglądów na temat historii Ameryki), następnie zaś jako „zburzenie domu i budowę na

<sup>96</sup> Case 1991, s. 665.

<sup>97</sup> Metafora dzięki dwutematycznej budowie, gdy o właściwościach jednego przedmiotu czy zjawiska orzeka się za pośrednictwem przedmiotu lub zjawiska innego, ujawnia nowe, niejednokrotnie bardzo subiektywne wyobrażenia o istocie metaforyzowanego elementu. Aby mógł zaistnieć proces metaforyczny, użytkownik języka musi dostrzec podobieństwo między dwoma różnymi elementami.

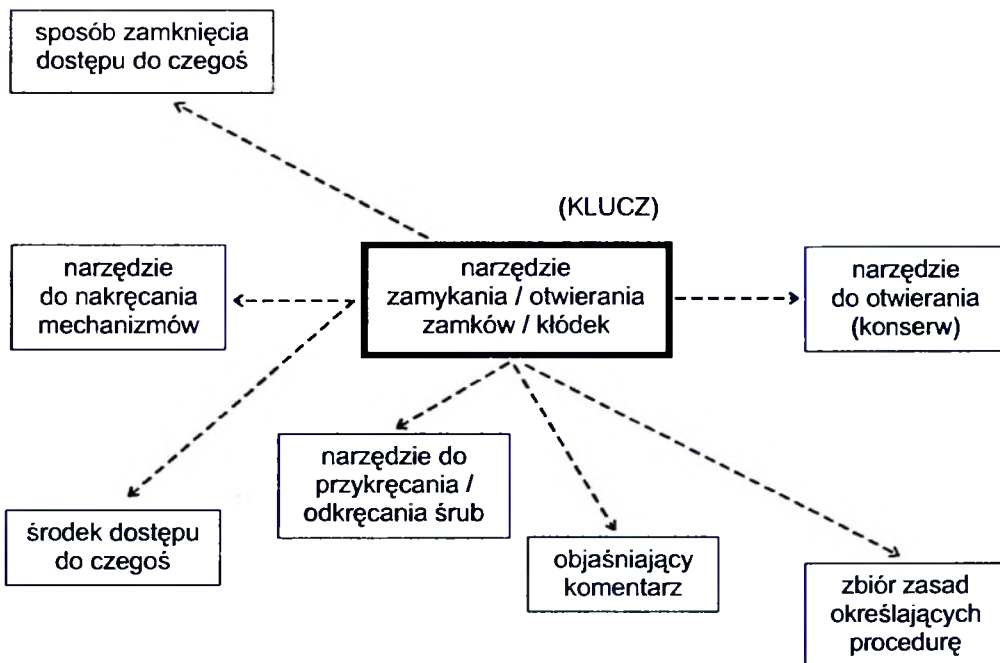
nowo” (tzn. przedstawienie całkowicie nowego spojrzenia). Ten sam historyk opisał nieśmiało próby zrewidowania historii jako „lukrowanie” tradycyjnych poglądów, by były łatwiejsze do połknięcia dla krytyków. Inny respondent parafrazował podejście młodych historyków do robienia notatek, posługując się dwiema metaforami, opisującymi jak historycy osiągają ważne cele: „Kiedy masz już dwa pudełka po butach wypełnione fiskami, powinieneś przystąpić do pisania swej dysertacji”. Ten szczególnie przypadek można uznać za przykład synekdochy, czyli figury stylistycznej, w której część zostaje uznana za całość: zbiór pomysłów równa się dysertacji.

Posługiwanie się metaforą przez przebadanych historyków wskazuje na użyteczność tej kategorii w wyrażaniu pomysłów na różnych poziomach abstrahowania. Z badań nad metaforą i metonimią w językach naturalnych prowadzonych przez Lakoffa i jego współpracowników rozwinęła się jedna<sup>98</sup> z dwóch koncepcji teoretycznych dotyczących związku struktur pojęciowych z językiem, znana pod nazwą modelu kategorii radialnej. Elżbieta Tabakowska<sup>99</sup> objaśnia istotę modelu kategorii radialnej na przykładzie polskiego leksemu *klucz*.

Kategoria radialna jest kategorią złożoną z elementów centralnych (prototypowych) i peryferyjnych (nieprototypowych). „Elementem centralnym kategorii jest element zaznaczony grubą ramką: <<narzędzie służące do zamykania i otwierania zamków i kłódek>>. Jest to znaczenie odpowiadające prototypowemu (dziś) rozumieniu polskiego leksemu *klucz*, czyli <<podstawowe>> w sensie słownikowym. Zauważmy, że takie znaczenie jest niedookreślone (schematyczne) – np. co do wielkości, szczegółów kształtu, materiału, z którego przedmiot jest zrobiony itp. W celu uszczegółowienia (ang. *elaboration, instantiation*) schematu, słownik na ogół podaje dodatkowe informacje, np. <<K. od biurka, szafy, szuflady, mieszkania>>. Pozostałe podane na rys. 1 znaczenia to odejścia od prototypu, czyli rozszerzenia (ang. *extensions*), włączone do kategorii na zasadzie podobieństwa do prototypu. Nie jest to jednak podobieństwo oznaczające, że wszystkie elementy mają wszystkie cechy definicyjne (jak w kategoryzacji przez podanie warunków koniecznych i wystarczających), lecz tzw. podobieństwo rodzinne. (...) <<Narzędzie do nakręcania mechanizmów>> (np. *klucz do [nakręcania] zegara*) ma taki sam kształt i wykonuje się nim analogiczną operację, natomiast nie służy do zamykania i otwierania. <<Narzędzie do przykręcania/odkręcania śrub>> (np. *klucz francuski*) jest już tylko podobne w kształcie do „prototypowego” klucza, podobnie jak wykonywany nim ruch obrotowy przypomina obrót klucza w zamku. <<Narzędzie do otwierania konserw>> (*klucz do konserw*) otwiera puszkę (ale jej nie zamyka!) przez ruch obrotowy, nie jest to jednak obrót w zamku. Wszystkie trzy przypadki ilustrują rozszerzenie zwane metoni-

<sup>98</sup> Drugą jest koncepcja modelu sieciowego stworzona przez Langackera. Oba modele okazały się niezwykle istotne dla teorii gramatyki kognitywnej.

<sup>99</sup> Tabakowska 1995, s. 46-50.



Rys. 1 Kategoria radialna KLUCZ<sup>100</sup>

micznym (*pars pro toto*), ponieważ podstawą do powstawania peryferyjnych znaczeń staje się pewna część znaczenia prototypu, która wysuwa się na pierwszy plan (kształt, funkcja itp.).

Pozostałe cztery rozszerzenia pokazane na rys. 1 to rozszerzenia metaforyczne. We wszystkich przypadkach relacja «otwierania» została przeniesiona z dziedziny zdarzeń fizycznych w dziedzinę zdarzeń abstrakcyjnych. Poszczególne znaczenia słowa *klucz* są uszczegółowieniami wspólnego ogólnego schematu: wiedza (np. o człowieku: *klucz do ludzkich serc*, o zapisie muzycznym: *klucz nutowy*, o rozwiązaniu zadania: *klucz do zbioru zadań*, o taksonomii: *klucz botaniczny*) jest zamknięta w pojemniku (ludzkim sercu, nutach na pięciolinii, w zadaniu, w systemie klasyfikacji roślin) i aby do niej dotrzeć, trzeba pojemnik otworzyć odpowiednim narzędziem. W kognitywnej teorii metafory schemat tego rodzaju nosi nazwę metafory pojęciowej (ang. *conceptual metaphor*). Taka schematyczna metafora realizuje się w postaci wielu szczegółowych obrazów (np. *dobre słowo otwiera ludzkie serca*), zwanych metaforami obrazowymi lub metaforami jednorazowego użycia (ang. *image lub one-shot metaphors*)<sup>101</sup>. Rozszerzenia znaczeń prototypowych zależą od czynników antropologicznych, kulturowych, socjologicznych itp., stając się pełnoprawnymi jednostkami języka, jeśli zostaną usankcjonowane konwencją językową.

<sup>100</sup> Tabakowska 1995, s. 47.

<sup>101</sup> Ibidem s. 47-49.

Nie można nie zauważyć, że terminologia stosowana przez amerykańskich historyków i względna waga poszczególnych terminów u różnych osób była odmienna. Wyjątek stanowiły jedynie pojęcia odnoszące się do czasu i przestrzeni. Droga do rozszerzania metafor „biurowych” historyków mogłaby prowadzić raczej poprzez aspekty chronologiczne i geograficzne niż standaryzację słownictwa dla danych tematów. Ewentualnym mechanizmem do bezpośredniego manipulowania zarówno przypisywaniem kategorii, jak i odszukiwaniem ich, mogłaby być ruchoma skala czasowa z etykietami dla okresów zapożyczonymi spośród terminów powszechnie używanych przez historyków. Podejście geograficzne może posługiwać się stylizowaną mapą. W przypadku badania Case’a byłyby to mapa Ameryki Północnej, z której można by „wybierać” stany (np. Wirginia), krainy geograficzne (np. New England), regiony historyczne (np. Trzynaście Kolonii).

Wyniki badań Case’a skłaniają też do pewnej refleksji na temat języków i systemów informacyjno-wyszukiwawczych, a zwłaszcza ich interfejsów. Należałoby rozważyć, czy w zapamiętywaniu procedur systemowych nie mogłyby pomagać użytkownikom metafory systemowe tworzone i kodowane w pamięci na kształt najbliższego otoczenia. Jeżeli pewne spostrzeżenia, myśli czy idee są osadzone w pamięci, to może należałoby je odwzorowywać także wizualnie i przestrzennie w interfejsie użytkownika. „Użytkowanie” stałoby się wówczas poruszaniem się po przestrzeniach informacji w poszukiwaniu fizycznego miejsca jej przechowywania. Nadal potrzeba też więcej badań kategorii abstrakcyjnych. Obszar kategorii abstrakcyjnych jest potencjałem dla badań zarówno w nauce o informacji, psychologii kognitywnej jak i w innych powiązanych naukach.





## KATEGORYZACJA: SELEKTYWNY PRZEGLĄD BADAŃ

Pojęcie kategoryzacji jest wieloznaczne. Najczęściej *kategoryzacją* nazywa się albo proces wyodrębniania obiektów i ich zbiorów traktowanych z pewnego punktu widzenia jako jednorodne, czyli jako proces ustanawiania kategorii, albo proces przypisywania okazów (przedmiotów) do już istniejących kategorii. W drugim przypadku używa się też określenia *kategoryzowanie*. Kategoryzacja bywa też rozumiana jako system kategorii. Termin kategoria z kolei najczęściej jest używany w znaczeniach:

1) klasa obiektów charakteryzujących się wspólnymi właściwościami wziętymi za podstawę kategoryzacji; czasami formułowany jest dodatkowy warunek odpowiedniej ogólności klasy aspirującej do miana kategorii jako odróżniający ją od klasy w ogóle;

2) podstawowe pojęcie danej nauki, np.: do podstawowych kategorii w etyce, obecnych w różnych systemach, choć często odmiennie interpretowanych, należą takie pojęcia jak: wartość, ideał, dobro, zło, norma, obowiązek, sankcja, wzór osobowy, szczęście, odpowiedzialność<sup>1</sup>;

3) zasadnicze sposoby bytowania odpowiadające podstawowym (ze względu na strukturę języka starogreckiego) rodzajom orzecznika przy najogólniejszej klasyfikacji bytu; najwyższy rodzaj, wewnątrz którego istnieje hierarchia rodzajów mniej ogólnych i ich gatunków, odpowiadający podstawowym aspektom, ze względu na które można rozważać każdy przedmiot i każdą dziedzinę bytu. Spośród wielu rozmaitych systemów kategorii ontologicznych<sup>2</sup>

<sup>1</sup> Kategorie etyki cytowane za: Mały słownik 1994, s. 108-109.

<sup>2</sup> We współczesnej ontologii najczęściej bierze się pod uwagę następujące kategorie: rzecz, zdarzenie (lub proces), cecha i relacja. Mill wyróżniał jako kategorie: uczucia, dusze, ciała, stosunki, Wundt – rzeczy, cechy, stany, stosunki. Reizm za podstawową (a raczej jedyną realną) kategorię uważał rzecz; ewentyzm – zdarzenie; materializm dialektyczny rozpatrywał pary przeciwstawnych kategorii ontologicznych, nie ustalając ich liczby: byt – świadomość, treść – forma, istota – zjawisko, ilość – jakość, możliwość – rzeczywistość, konieczność – przypadek itd. Filozofowie opierający ontologię na teorii mnogości traktują jako kategorie ontologiczne indywidua (clemen-

i metafizycznych, najstarszy i najbardziej znany stanowią kategorie<sup>3</sup> Arystotelesa;

4) aprioryczne pojęcia czystego intelektu w rozumieniu Kanta, odpowiadające podstawowym funkcjom logicznym rozumu i odnoszące się a priori do przedmiotów naoczności w ogóle; sposób myślenia, przejawiający się m. in. w ocenach i preferencjach, np.: myślenie kategoriami religijnymi, estetycznymi itp. Kant wyróżnił dwanaście pojęć czystego rozumu podzielonych na cztery klasy: ilość (jedność, wielość, całość), jakość (twierdzenie, przeczenie, ograniczenie), stosunek (substancja, przyczyna, wspólność), modalność (możliwość, istnienie, konieczność). W tym znaczeniu termin kategoria jest zatem skróconą nazwą kategorii apriorycznych.

Te różne sposoby rozumienia terminu i, co za tym idzie, różne aspekty badania kategoryzacji przenikają się wzajemnie i na ogół muszą być rozpatrywane łącznie, co staramy się czynić w całej rozprawie.

*Kategoryzacja* wywodzi się od Arystotelesa, który tym terminem nazwał klasy pojęciowo-ontologiczne wyróżnione na podstawie orzeczników<sup>4</sup>. O to, czy kategoria oznacza klasę rzeczy, znaczeń czy niezależnych bytów, przez wieki ciągnął się spór filozoficzny, implikujący pytania o naturę znaku językowego. Spór ten był szczególnie intensywny w średniowieczu i obecnie jest znany jako spór o tzw. uniwersalia. We współczesnym językoznawstwie mocno akcentuje się swoistość ujmowania świata przez poszczególne języki. Każdy język dysponuje właściwą sobie siatką semantyczną, tworzącą specyficzny dla danego języka obraz świata. Tylko część tej siatki to kategorie Arystotelesowskie. Swoistość struktur pojęciowych poszczególnych języków przejawia się przede wszystkim w odmiennej segmentacji świata za pomocą słownictwa<sup>5</sup>. Na przykład, jednemu polskiemu słowu *mgła* odpowiadają w języku angielskim

---

ty) i ich zbiory, a właściwie całą hierarchię zbiorów. Ci zaś, którzy opierają ontologię na ogólnej teorii systemów, uważają za najważniejsze kategorie elementy i system, a właściwie hierarchię systemów. Z kolei filozofowie skupiający się na filozofii człowieka (egzystencjalizm, personalizm) traktują zwykle jako podstawowe i przeciwstawne dwie kategorie: rzecz i osobę, z których pierwsza istnieje tylko „w sobie”, druga zaś – „dla siebie”.

<sup>3</sup> Tutaj termin kategoria jest skróconą nazwą kategorii ontologicznych, będących podstawowymi sposobami istnienia bytu i orzekania o nim.

<sup>4</sup> *Katēgorein* [gr.] – pierwotnie ‘oskarżać’, później ‘twierdzić, orzekać’.

<sup>5</sup> Język ze swej natury jest skonwencjonalizowanym systemem symboli i w tym sensie można mówić, że narzuca użytkownikowi pewien system kategorii. Konwencjonalizacja nie musi jednak implikować arbitralności. Kategorie zakodowane w języku są w różnym stopniu determinowane przez czynniki takie jak nieskończoność rzeczywistości; sposoby, w jakie ludzie w danej kulturze wchodzi w interakcje ze światem; natura procesów kognitywnych, w trakcie których formowane są pojęcia; indywidualne, osobnicze interpretacje rzeczywistości itd. Przekonanie, że kategoryzacja świata nie ma charakteru obiektywnego, ale jest pewną interpretacją rzeczywistości narzuconą przez umysł poznający, który ustalając podobieństwa między obiektami, kategoryzuje świat subiektywnie, w sposób nieostry, na przykład opierając się na zjawiskach typowych i uznając je za reprezentatywne dla całej kategorii, jest podstawą podejścia kognitywnego. Kategorie zakodowane w jednym języku przeważnie nie pozostają w jedno-jednoznacznym stosunku do kategorii w innym języku lub innym podsystemie tego samego języka.

skim trzy klasy semantyczne oznaczane odrębnymi słowami: *mist* (mgła przeciętnie gęsta), *fog* (mgła gęstsza niż przeciętna), *haze* (mgła rzadsza niż przeciętna) i odwrotnie to, co Anglik nazwie jednym wyrazem *cloud*, Polak podzieli na chmurę i obłok; jednemu polskiemu wyrazowi *śnieg* w języku szwedzkim odpowiada co najmniej pięć wyrazów: *snöglopp*, *snöyra*, *modd*, *snösörja* i *kram* oraz wiele więcej w języku Eskimosów. Liczne są także odmienności w kategoriach gramatycznych, na przykład różnica między kategorią rodzaju w języku polskim i angielskim, w którym to odróżnienie występuje jedynie w nazwach istot żywych<sup>6</sup>.

Kiedy używamy słowa *pies* w odniesieniu do dwóch różnych zwierząt lub gdy opisujemy dwa różne doznania wzrokowe koloru tym samym wyrazem *czerwony*, dokonujemy kategoryzacji. Istotą kategoryzacji jest, jak piszą George Lakoff i Mark Johnson, porządkowanie świata na podstawie wybranych cech przy braku jednolitego kryterium ich doboru. „Kategoryzacja jest naturalnym sposobem rozpoznawania rodzaju rzeczy lub doświadczenia poprzez uwypuklenie pewnych właściwości, pomniejszenie innych i ukrywanie dalszych. [...] Wygłaszając jakieś stwierdzenie, dokonujemy jednocześnie wyboru kategorii [sic! – przypis JW], gdyż mamy pewne powody, aby skupić się na jednych właściwościach, a pomniejszyć inne (...), podczas gdy klasyfikacja dokonywana jest na podstawie wszystkich znanych cech, według jednolitego kryterium”<sup>7</sup>.

Kategoryzacja często utożsamiana jest z klasyfikacją. Synonimicznie traktuje te terminy na przykład Jolanta Maćkiewicz, gdy pisze „kategoryzacja (klasyfikacja) to termin, którym posługuje się zarówno językoznawstwo, jak i psychologia czy logika”<sup>8</sup>. Maćkiewicz proponuje wprowadzenie terminów *kategoria naukowa* – *kategoria nienaukowa* w miejsce *kategorii logicznej* (*klasycznej*) i *kategorii naturalnej*. „Rozróżnienie: kategoria logiczna (*klasyczna*) – kategoria naturalna dałoby się wobec tego zastąpić przez rozróżnienie: kategoria naukowa – kategoria nienaukowa. Sposób tworzenia kategorii logicznych odpowiada naukowemu dążeniu do ostrości i jednoznaczności używanych pojęć. Struktura kategorii naturalnych wiąże się z nienaukowym porządkowaniem świata na zasadzie podobieństwa do stereotypowych przykładów i kontrprzykładów, z kompleksowym charakterem poznania potocznego i jego antropocentryzmem”<sup>9</sup>.

Ale sam język jest także częścią świata. Mówiąc o rzeczownikach, czasownikach, fonemach, zdaniach itd., dokonuje się pewnej kategoryzacji. W pracy oba te wątki (przedmiotowy i metapredmiotowy) będą się splecać. Tak jak językoznawcy dokonują kategoryzacji obiektów językowych języków naturalnych, tak będziemy się starać dokonać kategoryzacji obiektów

<sup>6</sup> Przykłady przytoczone w tym paragrafie zostały przejęte z pracy: Maćkiewicz 1983.

<sup>7</sup> Lakoff 1988, s. 190-191. Cyt. za: Habrajska 1996, s. 223.

<sup>8</sup> Maćkiewicz 1990, s. 51.

<sup>9</sup> Ibidem s. 55.

językowych jiw i obiektów świata opisywanego za pomocą jiw. Jiw z racji swej metainformacyjnej natury, o czym pisaliśmy w rozdziale pierwszym, stwarzają dodatkowe problemy do rozwiązania. Jeśli założyć, że istnieje paralela pomiędzy kategoryzacją obiektów języka i dokonującą się za pośrednictwem języka kategoryzacją rzeczywistości pozajęzykowej, to powinien też istnieć analogiczny związek między obiektami jiw i obiektami odwzorowywanej przez nie rzeczywistości. I odwrotnie, im więcej będziemy wiedzieć o rzeczywistości pozajęzykowej jiw, tym lepiej będziemy mogli orzekać o strukturze samego języka informacyjno-wyszukiwawczego.

W językoznawstwie termin kategoria najczęściej jest rozumiany jako zbiór elementów, które pełnią tę samą funkcję w systemie języka, występują w identycznych kontekstach języka (mają taką samą dystrybucję). Najczęściej wyróżnia się kategorie: gramatyczne, syntaktyczne, fonologiczne (przede wszystkim fonemy i ich klasy), morfologiczne (fleksyjne i słowotwórcze), semantyczne i leksykalne<sup>10</sup>. Oprócz uniwersalnych kategorii językowych, tzn. występujących we wszystkich językach (np. podstawowe kategorie fonologiczne jak samogłoski i spółgłoski), istnieją też kategorie właściwe dla niektórych tylko języków (np. kategoria przypadku).

Kategoria gramatyczna to taka kategoria językowa, której elementy spełniają tę samą funkcję językową (pozostając w opozycji formalno-funkcyjnej do innych kategorii spełniających inne funkcje) i posiadają formalny wykładnik językowy przynależności do danej kategorii, na przykład w języku polskim kategoria liczby, gdzie formalnym wykładnikiem są końcówki czy kategoria osoby czasownika, której wykładnikami są także końcówki. Kategorie gramatyczne to pewne pojęcia, a ściślej przeciwstawienia pojęciowe, które w danym języku wyrażane są w sposób regularny (tzn. za pomocą określonych form gramatycznych) oraz konieczny (tzn. mówiący zobowiązani są do przekazania określonej wartości danej kategorii). Często kategorię wyznacza tylko jedna z funkcji współtworzących kategorię gramatyczną w powyższym znaczeniu, na przykład: kategoria pierwszej osoby, kategoria liczby pojedynczej itd. Niektórzy językoznawcy „rozdzielają kategorie gramatyczne pierwszego stopnia, zwane również kategoriami syntaktycznymi, do których zalicza się składniki pełniące określoną funkcję w zdaniu (fraza nominalna, fraza werbalna lub podmiot, orzeczenie, przydawka, dopełnienie) oraz kategorie gramatyczne drugiego stopnia, które pokrywają się z tradycyjnym pojęciem części mowy (rzeczownik, przymiotnik, czasownik itd.). Niekiedy, zwłaszcza w gramatyce generatywnej, te ostatnie obejmuje się pojęciem kategorii leksykalnych”<sup>11</sup>. Na przykładzie kategorii gramatycznych widać, że ka-

---

<sup>10</sup> Wyliczone kategorie nie tworzą klasyfikacji, są jedynie zbiorem najczęściej wyróżnianych kategorii. Na przykład kategorie syntaktyczne i morfologiczne bywają czasami wyróżniane jako podkategorie kategorii gramatycznej, zaś kategorie semantyczne w pewnych interpretacjach są traktowane jako kategorie syntaktyczne.

<sup>11</sup> Encyklopedia 1993a, s. 261.

tegorie mogą pozostawać względem siebie w stosunkach hierarchicznych. Znaczenia leksemów wyrażane za pomocą ich form gramatycznych mogą odсылać do pewnych zjawisk rzeczywistości pozajęzykowej i wówczas mówi się o funkcji semantycznej (nominatywnej) kategorii gramatycznych; mogą także sygnalizować jedynie relacje składniowe, wówczas mówi się o ich funkcji syntaktycznej. Ta sama kategoria gramatyczna w odniesieniu do różnych klas leksemów może pełnić różne funkcje.

Kategoria syntaktyczna, jak wcześniej wspomniano, to zbiór wyrażen mających jednakową wartość składniową, czyli charakteryzujących się jednakowymi potencjalnymi zastosowaniami składniowymi. Podział wyrażen na kategorie syntaktyczne stosuje się zarówno w logice, jak i w językoznawstwie, przy czym podziały te nie są równoważne. Logiczne podziały na kategorie syntaktyczne uwzględniają w znacznym stopniu semantyczne właściwości wyrażen i z tego powodu kategorie syntaktyczne są też określane mianem kategorii semantycznych. Ogólnie można stwierdzić, że w ramach logicznej kategoryzacji te same wyrażenia języków naturalnych musiałyby być równocześnie zaliczane do różnych kategorii oraz że nawet uproszczony podział na kategorie syntaktyczne stwarza znaczne trudności. Stwierdzają to sami logicy, którzy przyczyny tego stanu rzeczy upatrują w wielości funkcjonalnej wyrażen języka naturalnego i logiczną teorię kategorii syntaktycznych uznają dla tych języków za pewną idealizację.

Podział na kategorie syntaktyczne w teorii składni w językoznawstwie jest prostszy niż postulowany przez logiczną teorię kategorii syntaktycznych. Kategorie syntaktyczne są tu paralelne do kategorii semantycznych, abstrahując bowiem od formalno-funkcjonalnych właściwości wyrażen. Natomiast w teorii logicznej paralelizm ten jest zaciemniony przez fakt uwzględniania przy kategoryzacji, obok sensów, odmiennych formalnych właściwości wyrażen, np. różnic między czasownikami a rzeczownikami, które odzwierciedlają pewne różnice formalno-funkcjonalne, np. to, że w przeciwieństwie do czasownika rzeczownik w języku polskim nie może być funktorem zdaniotwórczym.

Na nasz użytek przyjmujemy, że kategorią syntaktyczną, zwaną też składniową lub syntagmatyczną, jest kategoria językowa, grupująca wyrażenia o takiej samej dystrybucji zdaniowej, czyli mogące występować w zdaniu w takim samym otoczeniu (kontekście). W formalnym opisie języka kategorie syntaktyczne wyróżnia się na wszystkich poziomach struktury języka. W zależności od tego, czy kategorie muszą być reprezentowane w zdaniu, czy też nie, dzieli się je na kategorie obligatoryjne i fakultatywne; w zależności od tego, czy mogą samodzielnie tworzyć zdania, czy też nie – na kategorie wyrażen autosyntaktycznych i wyrażen synsyntaktycznych; w zależności od funkcji konstytuujących rodzaj wyrażenia złożonego wyróżnia się np. kategorię wyrażen zdaniotwórczych i kategorię wyrażen niezdaniotwórczych.

„W gramatyce tradycyjnej za kategorie semantyczne uważane były klasy znane jako części mowy, w odróżnieniu od części zdania traktowanych jako klasy syntaktyczne. Wyróżniano je bowiem m. in. na podstawie kryteriów semantycznych. A więc np. rzeczowniki w klasycznej definicji semantycznej miały oznaczać przedmioty, przymiotniki – własności, czasowniki – czynności i stany, liczebniki – ilości itd. Ze względu na rozbieżności między tymi definicjami a rzeczywistym stanem rzeczy ratowano sytuację tworząc pojęcia typu «uprzedmiotowione czynności i stany» lub «czynności ujęte jako przedmioty» itd. Uzasadniano w ten sposób poprawność definicji kategorii semantycznej rzeczownika, w której obok nazw przedmiotów znajdują się i nazwy czynności, i jakości, i stanów i in. Części mowy przypominają zatem kategorie ontologiczne Arystotelesa, które też musiały być wyróżniane na podstawie odrębności wyrazów języka greckiego. W innym ujęciu przez kategorie semantyczne rozumie się klasy pojęciowe wyrazów, znane jako pola semantyczne lub pola pojęciowe. Wyrazy należące do poszczególnych pól semantycznych łączą pewne wspólne wartości semantyczne. Wyróżniono m. in. takie kategorie semantyczne, jak nazwy pokrewieństwa, nazwy barw, nazwy istot żywych, nazwy narzędzi, nazwy zawodów i in.<sup>12</sup> To drugie rozumienie kategorii semantycznej (jako klasy wyrazów wykazujących wspólne cechy semantyczne) bywa też określane mianem kategorii leksykalnej. Niekiedy różnicuje się znaczenia terminów kategoria semantyczna i kategoria leksykalna, przez kategorię semantyczną rozumiejąc klasę form językowych, które w swojej strukturze formalnej mają wykładniki przynależności do kategorii w postaci morfemów gramatycznych z funkcją znaczeniową. W pracy będziemy się posługiwać terminem kategoria leksykalna w znaczeniu klasa wyrażen (elementów systemu leksykalnego) wykazujących wspólne cechy znaczeniowe, pełniących pewną wspólną funkcję semantyczną.

Obecnie podejmiemy próbę przedstawienia wybranych koncepcji kategoryzacji, począwszy od poglądów Arystotelesa na kategorie, poprzez klasyczne<sup>13</sup> ujęcie kategoryzacji w językoznawstwie, a właściwie prawie wyłącznie w semantyce językoznawczej (zagadnienia gramatyki języków naturalnych nie miały do tej pory prawie żadnego wpływu na teorię i praktykę języków informacyjno-wyszukiwawczych), Wittgensteina koncepcję podobieństwa rodzinnego wyrazów, do kognitywnych kategorii prototypowych. Przedmiotem rozprawy nie jest jednak zagadnienie kategoryzacji w ogóle, kategoryzacji językowej czy nawet kategoryzacji leksykalnej, a jedynie kategoryzacji w językach

---

<sup>12</sup> Encyklopedia 1993a, s. 263.

<sup>13</sup> Użycie określenia *klasyczne* może się wydawać nadużyciem, jeśli wziąć pod uwagę z jednej strony fakt, że problemy semantyczne znalazły się w polu zainteresowania współczesnego językoznawstwa stosunkowo późno, bo dopiero w latach sześćdziesiątych obecnego stulecia, z drugiej zaś – wicość prac teoretycznych i szczegółowych z zakresu semantyki trudno poddającą się syntezie. Zastosowanie tego określenia jest jednak świadome i podyktowane chęcią wyrazistego zaznaczenia współlistnienia obecnie dwóch zasadniczych kręgów badań językoznawczych – kognitywnego i nickognitywnego (nazwanego tu klasycznym).

informacyjno-wyszukiwawczych. Dlatego niniejszy rozdział ma charakter selektywny, przeglądowy i w pewnym stopniu upraszczający. Mając tego świadomość, jak również pamiętając o interdyscyplinarnym charakterze problematyki kategoryzacji (cytowanie konkretnych publikacji jako prac z dziedziny filozofii, antropologii, psychologii czy językoznawstwa jest nierzadko sprawą umowną), chcemy przedstawić i przybliżyć czytelnikowi przede wszystkim tło językowe i językoznawcze, na którym w kolejnym rozdziale poddamy analizie zjawisko kategoryzacji i jej realizacji w wybranych językach informacyjno-wyszukiwawczych, dla których relewantny i efektywny system kategorii ma ogromne, choć ciągle niedoceniane znaczenie. Zaczynamy od przedstawienia poglądów Arystotelesa jako prapoczątku wszelkich badań nad kategoryzacją.

## KATEGORIE ARYSTOTELESOWSKIE

Ontologiczna kategoryzacja bytów Arystotelesa do dziś jest zarówno obiektem rozważań naukowych, jak i punktem odniesienia wielu podziałów kategoryalnych w językach naturalnych i sztucznych w tym w językach informacyjno-wyszukiwawczych.

Arystoteles głosił, że byt samoistny, czyli substancjalny, mają tylko jednostkowe rzeczy oraz że istotą poznania jest genetyczny empiryzm i koncepcja zmysłowego pochodzenia wszelkiej wiedzy, nierozzerwalnie związanej z obiektywną rzeczywistością. Z platonizmu zachował przekonanie, że prawdziwa wiedza dotyczy tylko tego, co ogólne i wyraża się w pojęciach ogólnych. Starając się rozwiązać zagadnienie stosunku tego, co ogólne, do tego, co jednostkowe, stosunku pojęć do rzeczywistości, Arystoteles zajął stanowisko umiarkowanego realizmu pojęciowego, będącego epistemologiczną konsekwencją teorii materii i formy, przyjętej przez niego za podstawę dociekań ontologicznych. Odrzucił zarówno twierdzenie Platona, że substancją są idee (formy), jak i tezę Demokryta, że substancją jest materia, twierdząc, że każda substancja, tj. rzecz, jest zawsze bytem złożonym z dwóch nierozzerwalnie związanych składników-zasad: materii, pojętej jako bierne tworzywo, stanowiące podłoże wszelkiego stawania się, i wszelkiej egzystencji konkretnej, oraz formy, pojętej jako czynnik aktywny, kształtujący z materii rzecz danego rodzaju. Materia i forma stanowiły według Arystotelesa dwa aspekty tego samego bytu. Forma mogła istnieć tylko w połączeniu z materią jako inteligibilny<sup>14</sup> składnik rzeczy. W swoich rozważaniach Arystoteles położył zasadniczy nacisk na formę jako aktywny, rozumowy pierwiastek bytu. Forma była tym, co w naturze daje się ująć rozumowo, co stanowi przedmiot wiedzy ogólnej, co konstytuuje istotę rzeczy.

---

<sup>14</sup> *Intelligibile* – termin filozoficzny, oznaczający to, co poznawalne tylko rozumowo a nie zmysłowo i co stanowi przedmiot wiedzy abstrakcyjnej.



Przez istotę rzeczy Arystoteles rozumiał zespół cech trwałych bytu, wyznaczających jego cechy swoiste, tj. stanowiące o tym, że byt ów jest taki właśnie, jaki jest. Można przyjąć<sup>15</sup>, że istota rzeczy stanowi powszechnik (pojęcie ogólne), a więc to, co jest wspólne wszystkim przedmiotom nazywanym daną nazwą. Poszukiwanie istoty rzeczy doprowadziło do późniejszego średniowiecznego sporu o uniwersalia, w którym ścierały się co najmniej cztery stanowiska, występujące dalej w filozofii w różnych postaciach i pod różnymi nazwami. Były to:

„a) nominalizm: rzeczy mają charakter jednostkowy, a tylko nazwy są ogólne – uniwersalia *post rem*;

b) skrajny idealizm platoński: pojęcia ogólne (idee) mają byt niezależny (realizm platoński) – uniwersalia *ante rem*;

c) realizm umiarkowany arystotelesowski: uniwersalia (cechy ogólne) tkwią w rzeczach, a umysł ludzki je tylko odkrywa – *uniwersalia in re*;

d) konceptualizm: pojęcia ogólne tylko w umyśle – *uniwersalia in intellectu*”<sup>16</sup>.

Istota rzeczy jest fundamentem przydatności definicji realnych, od których wymaga się, aby podawały charakterystykę nie tylko jednoznaczną, ale i istotną (wskazującą cechy istotne definiowanego przedmiotu). U Arystotelesa nie ma rozróżnienia, które wprowadziła późniejsza logika, na definicję nominalną i definicję realną. Dla Arystotelesa definicja właściwa to nie wyjaśnienie znaczenia słowa, ale opis przedmiotu nazywanego tym słowem, opis, który oddaje istotę tego przedmiotu, jego ogólną naturę. Arystotelesowska koncepcja definicji została później sprecyzowana jako charakterystyka zakresu danej nazwy ogólnej za pomocą wskazania cech koniecznych i wystarczających, przysługujących wszystkim elementom zbioru, do którego dana nazwa się odnosi<sup>17</sup>.

Według Arystotelesa każda rzecz, każdy przedmiot jest „zbudowany” z czynników konstytutywnych (określających) i konsekwentnych (pochodnych), z zestawu elementów noszących i noszonych, czyli z czynników tworzących substancjalność, podmiotowość rzeczy i akcydentalność. Co zatem w rzeczy stanowić miałyby istotę, różną od właściwości akcydentalnych? Mieczysław Krąpiec wyraża pogląd, że intuicje co do opozycji konstytutywny ÷ konsekwentny mają charakter generalny i zostały „nadbudowane na oglądzie wytworów sztuki – rzemiosła, tudzież oglądu osobników występujących w przyrodzie: roślin i zwierząt. (...) w historii nauki (np. w Akademii pla-

---

<sup>15</sup> Arystotelesowski sposób do uniwersaliów nie było do końca jasne i swe klarowne rozwiązanie znalazło dopiero w *De ente et essentia* św. Tomasza z Akwinu. „Powszechniki, zdaniem św. Tomasza, są konstruktem intelektu z podstawami w rzeczy. Ogólność, schematyczność powszechnika (pojęcia ogólnego) jest wytworem umysłu poznającego i ujmującego niektóre zasadnicze (z jakiegoś punktu widzenia) cechy. Dzięki temu ujmowane indywidualia „mieszczą się” w zakresie tak wytworzonego powszechnika”. Cyt za: Krąpiec 1985, s. 79.

<sup>16</sup> Językowa 1996, s. 14.

<sup>17</sup> Tej klasycznej koncepcji definicji przeciwstawiają się definicje kognitywne (oparte na prototypie) zgodne z którymi granice klas wyróżnionych przez nazwy nie są ostre, a elementy w różnym stopniu przynależą do zakresu nazwy.

tońskiej) model poznawania istoty został zaczerpnięty z twórców geometrii, takich jak: punkt, linia, trójkąt, koło, kwadrat – a więc z konstrukcji idealnych, niezwykle prostych (niezłożonych), gdzie wyliczenie komponentów danej figury bywało w gruncie rzeczy bardzo łatwe. Odpowiednie wyliczenie komponentów danej figury geometrycznej stanowiło wyliczenie jej cech strukturalnych, konstytutywnych, podczas gdy inne sprawy, jak podłoże i związane z tym właściwości (zieloność, czerwien itp.) uchodziło za akcydensy w stosunku do rozumianej istotnościowo figury”<sup>18</sup>. W szkole arystotelesowskiej model poznania istoty rzeczy został z kolei związany z hierarchią klas naturalnych, obserwowaną w przyrodzie. Jednak „taka właśnie szeroko rozumiana hierarchia bytowa, jako «świadectwo» możliwości istotnościowego poznania przez intelekt struktur bytowych, gatunkowych, jest w gruncie rzeczy poznaniem bardzo pobieżnym, świadczącym o „słabości intelektualnej” człowieka, który najczęściej ujmuje tylko bardzo szerokie zarysy organizacji przyrody i nie może zbyt łatwo wskazać na czynnik gatunkujący w obrębie niemal oczywistych klas naturalnych. (...) Na pierwszy rzut oka odróżniam krowę od konia i owcy, ale nie ujmuje jakiegoś jednego czynnika gatunkującego, lecz ograniczam się do wyliczenia zespołu charakterystycznych cech danych grup zwierząt w obrębie uznanej hierarchii klas naturalnych”<sup>19</sup>.

Dla rozumienia rzeczywiście, a przez to formowania pojęcia bytu realnego, szczególnie doniosła była analiza sposobu orzekania (predykcji) ujawniająca właśnie istotowe i akcydentalne cechy strukturalne rzeczy. Warto w tym miejscu przypomnieć, że dzieło Arystotelesa *Kategorie* zajmowało się właśnie rodzajami orzeczników, a nie klasami bytów opisywanych za pomocą cech definicyjnych, co mogłoby znaczyć, że pierwotnie u Arystotelesa pojęcie kategorii było rozumiane nie w takim sensie, jakie się mu dziś przypisuje<sup>20</sup>. W Księdze pierwszej *Topik* Arystoteles dokonał podziału predykabiliów<sup>21</sup> na: definicje, własności, rodzaje i cechy przypadkowe. „Teraz z kolei trzeba określić rodzaje kategorii, w których występują owe cztery wymienione terminy. Jest ich dziesięć: istota, ilość, jakość, stosunek, miejsce, czas, pozycja, stan, czynność, bierność. Otóż cecha przypadkowa i rodzaj, i własność, i definicja znajdują się zawsze w jednej kategorii; wszystkie bowiem przesłanki, które powstały dzięki nim, oznaczają albo istotę jakiejś rzeczy, albo jej jakość, albo ilość, bądź inną z pozostałych kategorii”<sup>22</sup>. Komentatorzy Arystotelesa zwykle wyróżniali orzeczniki ze względu na to, że oznaczają one w samej rzeczy albo

<sup>18</sup> Krąpicc 1985, s. 184-186.

<sup>19</sup> Ibidem s. 188.

<sup>20</sup> Wprowadzić już u Arystotelesa kategorie zaczynają oznaczać także działy przedmiotów, czyli nabierają wtórnie znaczenia ontologicznego, jednak wyraźniejsze to jest u kontynuatorów Arystotelesa, przede wszystkim u Porfiriusza, który we wstępie do *Kategorii*, tzw. *Isagoga*, mówi o kategoriach bytów, szeregując je według malejącej ogólności w tak zwanym drzewie Porfiriusza

<sup>21</sup> Predykabilia – termin filozoficzny, oznaczający sposoby orzekania pojęć ogólnych o przedmiocie jednostkowym.

<sup>22</sup> Arystoteles 1990b, s. 350

istotę, albo coś niekoniecznego. Jeśli orzeczniki dotyczą istoty, to albo mają na względzie całą istotę i wówczas mówi się o gatunku, albo tylko jej część, która może być częścią potencjalną, i wówczas ma się do czynienia z rodzajem, albo część aktualną i wtedy mówi się o różnicy gatunkowej. Jeśli orzeczniki oznaczają coś, co do istoty rzeczy nie należy, ale jest z nią związane, to ów związek z istotą bywa albo koniecznościowy i wtedy jest to właściwość (cecha, która przysługując tylko temu gatunkowi, a nie innemu, przysługuje jednocześnie każdej jednostce tego gatunku i to w sposób nierozdzielny, np. u człowieka zdolność do płaczu), albo niekoniecznościowy i wtedy jest to przypadłość. Zarówno właściwości, jak i przypadłości w orzekaniu, są akcydensami w bytowaniu. Indywiduum nie jest orzecznikiem, ale podmiotem orzeczeń uniwersalnych. Samo indywiduum jest celem poznania poprzez różnego typu orzeczniki. Arystotelesowskie predykabilia opisujące świat zostały następnie przez gramatyków greckich zastosowane do języka jako typy części mowy, stając się początkiem tradycji kategorii gramatycznych.

Zespół predykabiliów stał się podstawą systematyzacji ontologicznej, albowiem „predykaty są też jakby istnościami różnego rodzaju, ustosunkowanymi do podmiotu na różne sposoby. Z racji tej kategorii dopuszczają również klasyfikację wszystkich rzeczy w zależności od sposobu ich istnienia. Są zatem ostatecznymi i nieredukowalnymi klasami, do których się dochodzi przez kolejne stawianie pytania „Czym dana rzecz jest?” aż do ostatecznego kresu”<sup>23</sup>. W rezulacie Arystoteles wyodrębnił 10 kategorii: substancja, ułożenie (położenie), mienie (stan), ilość, jakość i relacja, działanie i doznawanie działania (doznanie) oraz czas i przestrzeń (miejsce). „Każdą nie połączoną wypowiedź oznacza albo substancję, albo ilość, albo jakość, albo stosunek, albo doznawanie. Przykładem substancji jest – mówiąc ogólnie na przykład człowiek, koń; ilości – dwułokciowy, trzylokciowy; jakości – biały, gramatyczny; stosunku – podwójny, połowa, większy; miejsca – w Liceum, na rynku; czasu – wczoraj, ubiegłego roku; położenia – leży, siedzi; stanu – obuty, uzbrojony; działania – tnie, pali; doznawania – jest krajany, jest palony. Żadna z wymienionych wypowiedzi nie występuje sama w jakimkolwiek twierdzeniu; dopiero przez wzajemne ich połączenie powstaje twierdzenie i przeczenie; bo każde twierdzenie musi być – jak się przypuszcza – albo prawdziwe, albo fałszywe; natomiast wyrazy, które nie występują w żadnym połączeniu, nie są ani prawdziwe, ani fałszywe, na przykład: <<człowiek>>, <<biegnie>>, <<zwycięza>>”<sup>24</sup>.

Krąpiec tak objaśnia kategorie wymienione przez Arystotelesa: „Substancja i trzy rodzaje przypadłości: ilość, jakość i relacja, stanowią podstawę wyróżnienia i uznania innych właściwości-przypadłości bytu. Jeśli bowiem bytem w swej strukturze istotowej, koniecznej, jest substancja, tworząca sposób bytowania niezapodmiotowany, to bezpośrednio zapod-

<sup>23</sup> Leśniak 1990, s. 27.

<sup>24</sup> Arystoteles 1990a, s. 34.

miotowanym w substancji sposobem podzielnego bytowania jest ilość, a sposobem niepodzielnego bytowania jest jakość, doskonałą rzecz ze względu na niepodzielną formę bytową. Zapodmiotowaniem niebezpośrednim – «ze względu na» – jest relacja, która zawsze jest swoistym sposobem bytowania «między» jednym a drugim bytem lub jego częścią. Kategoria czasu – «quando» i przestrzeni («umiejscowienia») – «ubi – gdzie» wiąże się z ilością mierzalną – rozciągłościową statycznie lub będącą w ruchu (materia w ruchu). Zatem czas i przestrzeń w oczach Arystotelesa nie stanowią jakichś apriorycznych form bytowania materii, ale dotyczą bytu materialnego zilościowanego, różnie mierzalnego, a przez to w różny sposób dostępnego «użyciu» i «obróbce» tegoż materialnego bytu.

Dwie inne właściwości bytu: «actio – passio», działanie i doznawanie działania, pojawiają się człowiekowi jako właściwości użyteczne, skoro tylko przez wyzwolenie działania i zdolność recepcji (doznania trwałe i nietrwałe) można zużytkować, przekazać i powiększać jakości, niezbędne dla życia ludzkiego w jego najrozmaitszej postaci. Oczywiście nie znaczy to, by sprawa «działania – doznawania» sprowadzała się jedynie do zużytkowania jakości. Rozróżnienie tych jakości nabudowanych na substancji ma wartość teoretyczno-poznawczą, ale te właśnie wartości teoretyczne zostają zużytkowane przez człowieka w jego ludzkim zachowaniu się i to w pełnym kontekście bycia człowiekiem.

Dwie ostatnie kategorie «situs» i «habitus» – «ułożenie» i «mienie» – ukazują byt substancjalny, materialny (zilościowany) ze względu na występujące układy relacji ilościowych wewnętrznych i zewnętrznych. Byt bowiem zilościowany, zwłaszcza człowiek (ale i zwierzę) może sam zmieniać układy swego ciała i ustanawiać nowe relacje pomiędzy częściami swego organizmu, ze względu na działanie lub doznawanie. Ułożenie się ciała człowieka czy zwierzęcia jest ważne dla realnego zrozumienia samego bytu w danym czasie. Bardzo ważne są też relacje tegoż bytu materialnego do przedmiotów, które on «posiada», które są jego «mieniem». Uzbrojenie, ozdoby, przedmioty użytku ludzkiego notyfikują charakter bytu substancjalnego i stanowią niekiedy istotny czynnik oceny drugiego człowieka tak w walce, jak i w czasie pokoju. Sprawa «mienia» ludzkiego jest więc bardzo ważnym czynnikiem determinującym ludzkie bytowanie i działanie”<sup>25</sup>.

Wyróżnienie przez Arystotelesa 10 kategorii bytu było w filozofii chyba pierwszą próbą racjonalnego uporządkowania rzeczywistości, w której człowiek żyje i której jest także swoistą częścią. Znane wcześniej pięć Platónskich rodzajów form (byt, ruch, spoczynek, tożsamość, różnica) nie było w rozumieniu Arystotelesa kategoriami. „Na przykład byt. Arystoteles często podkreślał, że byt i jedność nie są kategoriami, ponieważ mogą być orzekane o wszystkim, a więc nie mieszczą się w żadnej z dziesięciu kategorialnych szufladek. To samo dotyczy tożsamości i różnicy. O każdej rzeczy można wszak powiedzieć,

<sup>25</sup> Krąpicc 1985, s. 203-204.

że jest tożsama ze sobą i że jest różna od każdej innej rzeczy. Z tej właśnie racji tych dwóch pojęć nigdy nie zalicza do kategorii. Nigdy by też nie uznał za kategorie ani ruchu ani spoczynku”<sup>26</sup>.

Pojęcie kategorii pojawiło się jeszcze raz w dziejach myśli filozoficznej w filozofii Kanta i neokantystów w postaci kategorii umysłu ludzkiego, warunkujących poznanie świata przez człowieka. Kategorie kantowskie nie są jednak własnością świata, ale zostają na świat „nałożone” przez umysł poznający. System szczegółowych kategorii w filozofii Kanta oparty był na właściwościach sądów. Najistotniejsza była jednak teza główna, głosząca, że poznanie świata zależy od kategorii pojęciowych właściwych umysłowi ludzkiemu. Porządkują one świat poznawany, umożliwiając człowiekowi nawet zwykłą percepcję, która „rozproszona” jest między różne zmysły – umysł pełni funkcję scalającą. Wydaje się, że pogląd ten jest bliski temu, co głoszą współcześni kognitywiści, choć na ogół w swoich tekstach nie nawiązują jawnie do filozofii Kanta.

Uporządkowanie rzeczywistości według dziesięciu kategorii mogło się dokonać dzięki językowi używanemu jako manifestacja stanów poznawczych komunikowanych sobie samemu i innym. Do kategorii Arystotelesa nawiązuje też cała późniejsza teoria mnogości, opisująca świat za pomocą pojęć przedmiotu, własności i relacji oraz pojęć pochodnych – zbiorów przedmiotów i relacji między zbiorami, zbiorów zbiorów i typów relacji między zbiorami wyższego rzędu, w wyniku czego powstaje bardzo rozbudowana struktura pojęciowa. Teoria mnogości stała się z kolei podstawą wyszukiwania informacji w wielu językach informacyjno-wyszukiwawczych, w tym we wszystkich jiw wykorzystujących regułę indeksowania współrzędnego. Kategorie ontologiczne Arystotelesa stały się modelem kategorialnym w wielu językach sztucznych, między innymi w językach informacyjno-wyszukiwawczych, a także punktem odniesienia klasyfikacji wyrażen w językach naturalnych. Przykładem może tu być kategoryzacja zastosowana w leksykonie Rogeta<sup>27</sup>, gdzie wyróżniono sześć kategorii: STOSUNKI ABSTRAKCYJNE (ang. ABSTRACT RELATIONS), PRZESTRZEŃ (ang. SPACE), MATERIA (ang. MATTER), INTELEKT (ang. INTELLECT), WOLA (ang. VOLITION), WŁADZE UCZUCIOWE i MORALNE (ang. AFFECTIONS).

Podsumowując, kategorie i kategoryzacja w ujęciu klasycznym, Arystotelesowskim, oparte są na trzech następujących przekonaniach:

(1) przynależność do kategorii jest określana poprzez koniunkcję cech koniecznych i wystarczających. Cechy te są binarne, tj. przedmiot może albo posiadać daną cechę, albo jej nie posiadać, co równa się albo przynależeniu do danej kategorii, albo nieprzynależeniu. W konsekwencji poszczególne kategorie mają ostre granice określone przez cechy definiujące;

---

<sup>26</sup> Leśniak 1990, s. 28.

<sup>27</sup> Roget 1966.

(2) kategorie tworzą układy hierarchiczne, w których kategoria niższego rzędu zawiera się w kategorii rzędu wyższego, a co za tym idzie, wszystkie elementy kategorii wyższego rzędu i elementy kategorii rzędu niższego posiadają wspólny zbiór cech identyczny ze zbiorem cech koniecznych i wystarczających elementów kategorii rzędu wyższego, zaś samą przynależność elementów do kategorii określają logiczne zasady sprzeczności i wyłączonego środka: nie można jednocześnie być i nie być elementem danej kategorii, trzeba natomiast albo nim być, albo nim nie być;

(3) wszystkie elementy kategorii są równe sobie (mają równy status), tzn. żaden element klasycznej kategorii nie jest jej „lepszym” czy „gorszym” reprezentantem od jakiegokolwiek innego elementu.

Arystotelesowska teoria kategorii jest odbiciem dążenia umysłu ludzkiego do uporządkowania świata. W rzeczywistości jednak większość zjawisk ma charakter skalarny. Językowe określenia kategorii rzekomo antonimicznych bywają złudne – duży kot (jak na kota) okazuje się mały w porównaniu z tygrysem. Od czasów Arystotelesa kategoryzację definiuje się zakładając, że kategorie tworzone przez umysł ludzki (kategorie poznawcze) odpowiadają kategoriom istniejącym w otaczającym nas świecie, inaczej mówiąc, że kategorie (cechy ogólne) tkwią w rzeczach, a umysł ludzki je tylko odkrywa. Obserwacja uczy jednak, że często kategorie poznawcze nie pokrywają się z „prawdziwymi” kategoriami otaczającej nas rzeczywistości, dając obraz takiego świata, jaki widzimy i w jaki wierzymy, a nie takiego, jakim on jest w percepcji innych ludzi.

Próba rozwiązania tego podstawowego dylematu legła u podstaw niearystotelesowskich systemów kategoryzacji, które są zbudowane na odmiennych założeniach. Najbardziej popularna okazała się prototypowa koncepcja kategorii. Powstała ona na gruncie szeroko zakrojonych badań empirycznych z dziedziny psychologii rozwojowej, a za jej twórczynię uznaje się amerykańską uczoną Eleanor Rosch. Więcej informacji na ten temat w rozdziale *Kategorie prototypowe*.

## KATEGORYZACJA W SEMANTYCE JĘZYKOZNAWCZEJ

Kategoryzacja jest niezbywalnym atrybutem poznania. Nerozstrzygnięta pozostaje przy tym kwestia, czy to struktura języka wpływa na percepcję i poznanie świata, w tym kategoryzację, czy też poznanie determinuje fakty językowe. Nasuwa się tu analogia co do autonomicznego lub kognitywnego charakteru znaków językowych i językoznawstwa w ogóle. Przekonanie, że znaczenie formy językowej wypływa z samego systemu języka jest charakterystyczne dla paradygmatu strukturalistycznego. Interakcje użytkowników języka z rzeczywistością, postrzeganie i konceptualizowanie w świetle myśli

strukturalistycznej są zjawiskami o charakterze pozajęzykowym i jako takie nie mają wpływu na system języka, który pozostaje autonomiczny.

Autonomia języka nabrała innego znaczenia w koncepcji generatywno-transformacyjnej Noama Chomsky'ego. Język zaczął być postrzegany jako organiczna, choć autonomiczna, część istoty ludzkiej (tak jak „system wiedzy” rezydujący w ludzkim mózgu) składająca się z mechanizmów generowania i transformacji, na mocy których użytkownik języka tworzy i rozumie nowe zdania, tzn. takie, z którymi poprzednio się nie zetknął. Zdaniem Chomsky'ego gramatyka transformacyjno-generatywna jest modelem kompetencji językowej użytkowników danego języka. Nie tylko generuje nieskończony zbiór zdań tego języka, ale także opisuje ich budowę syntaktyczną i wyznacza interpretację semantyczną dla każdego wygenerowanego zdania zgodnie z intuicją językową jego użytkowników. Chomsky uważał kompetencję językową tylko za jeden z czynników, obok takich jak pamięć, stan psychofizyczny mówiącego w czasie wypowiedzianego zdania itd., wpływających na zachowania językowe użytkowników języka, które określał mianem performancji. W koncepcji Chomsky'ego językoznawstwo jest dyscypliną autonomiczną, a przedmiotem jego zainteresowania była jedynie kompetencja językowa użytkownika idealnego – pozbawionego cech i zachowań osobniczych. Ów najogólniej pojęty indywidualny czynnik ludzki nie jest zdaniem generatywistów elementem kompetencji językowej, lecz elementem wykonania, a więc nie stanowi obiektu badania językoznawczego<sup>28</sup>.

Hipoteza o autonomicznym charakterze języka i jego poszczególnych komponentów krytykowana jest między innymi przez reprezentantów językoznawstwa, które nazwano kognitywnym<sup>29</sup>. Uprawiający językoznawstwo kognitywne, między innymi Ronald Langacker i George Lakoff, których prace<sup>30</sup> wydają się być kamieniami milowymi tej dyscypliny, uważają, że nie jest uzasadnione mówienie o „zmyśle” językowym jako niezależnym od innych zmysłów, zdolności poznawczych, percepcji, pamięci, uwagi, interakcji społecznych, osobowości, doświadczenia itp. Inaczej mówiąc, podstawowym założeniem badawczym językoznawców kognitywnych, bez względu na istniejące między nimi różnice poglądów, jest nierozdzielanie ludzkich zdolności i możliwości czysto językowych od innych. Wielu językoznawców, jak Anna Wierzbicka, którzy nie deklarowali się jako kognitywiści, wniosło istotny wkład do szukania wyjaśnienia pewnych problemów języka poza „zmysłem” językowym i autonomicznym systemem samego języka. Również w językoznawstwie polskim pewni jego wybitni przedstawiciele (Adam Heinz, Zenon

---

<sup>28</sup> Warto w tym miejscu przypomnieć, że sam Chomsky, nawiązując do tradycji racjonalistycznych Leibniza i Kartezjusza i wyznaczając w swojej teorii istotną rolę idcom wrodzonym, zwykł ją nazywać mentalistyczną i umiejscowił jako gałąź psychologii.

<sup>29</sup> Podejście kognitywne, także w językoznawstwie, jest znacznie starsze niż sama nazwa językoznawstwo kognitywne.

<sup>30</sup> Langacker 1987-1991; Lakoff 1987.

Klemensiewicz, Jerzy Kuryłowicz) prowadzili badania, które choć oparte na zasadach klasycznego strukturalizmu, są bliskie także tezom językoznawstwa kognitywnego poprzez ich związki z kulturą i problematyką społeczną.

Kognitywiści za źródło, a zarazem rację istnienia języka, uznają uwarunkowane kulturowo, społecznie i osobniczo zdolności ludzkiego umysłu, dzięki którym obserwacja otaczającego świata ostatecznie prowadzi do zdobywania wiedzy o tym świecie. Kluczem do zrozumienia natury języka jest zatem zrozumienie natury procesów ludzkiego poznania. Zdaniem kognitywistów fonologia, słownictwo, morfologia, składnia, semantyka i pragmatyka stanowią jedno continuum, będące odbiciem procesów poznawczych. Językoznawstwo kognitywne, przynajmniej w wydaniu Langackera, opiera się na podstawach strukturalistycznych i jest konsekwentną kontynuacją osiągnięć klasycznego strukturalizmu. Opowiada się jednak przeciwko nadmiernemu formalizmowi, charakterystycznemu zwłaszcza dla językoznawstwa amerykańskiego, pozostającego pod silnym wpływem koncepcji transformacyjno-generatywnej oraz przeciwko pomijaniu w opisie języka danych wymykających się uogólnieniom.

Jeśli chodzi o negowanie poglądu, że znaczenia form językowych stanowią niezależnie istniejące jednostki, to zgodni są zarówno językoznawcy kognitywni, jak i strukturaliści. Zarówno dla kognitywistów, jak i strukturalistów znaczenia są uzależnione kontekstowo. Dla strukturalistów zależność kontekstowa jest kwestią relacji syntagmatycznych i paradygmatycznych między znakami w systemie językowym, a więc kontekst dla znaczenia jest wewnątrzjęzykowy. Dla kognitywistów kontekst jest zewnętrzny w stosunku do systemu językowego, znaczenia są strukturami kognitywnymi wtopionymi we wzorce wiedzy, przekonań itd. W przeciwieństwie do ujęcia strukturalistycznego, w kognitywizmie znaczenie ujmuje się jako z zasady niezależne od struktur zleksykalizowanych w poszczególnych językach. Znaczenie formy językowej można zrozumieć jedynie w kontekście innych struktur kognitywnych; w zasadzie nieważne (nierелеwantne) jest to, czy te inne struktury kognitywne są w języku komunikowane leksykalnie. Na przykład, co jest znaczeniem słowa *poniedziałek*? *Poniedziałek* można zdefiniować jedynie w kontekście ‘tydzień’; ktoś nie zaznajomiony z pojęciem siedmiodniowego tygodnia nie będzie dysponował podstawą do zrozumienia *poniedziałku*. Z kolei ‘tydzień’ musi być rozumiany w kategoriach powtarzającego się cyklu dzień – noc. Wreszcie, pojęcia ‘dzień’, ‘tydzień’ i ‘poniedziałek’ rozumiane są w odniesieniu do pojęcia czasu. Podobnie, w *górze* – *w dół*, *wysoko* – *nisko*, *podnosić się* – *spadać* można wyjaśnić jedynie poprzez trójwymiarową przestrzeń grawitacyjną; *na zewnątrz* – *wewnątrz*, *wchodzić* – *wychodzić* wymagają pojęcia trójwymiarowego pojemnika; *narodziny*, *starość* i *śmierć* presuponują wiedzę o cyklach życia, *ciężki* wymaga odwołania się do ‘waga’ itd. Można zatem powiedzieć za Langackerem, że siedmiodniowy tydzień jest domeną semantyczną, w odniesieniu do której rozumiany jest *poniedziałek*, zaś cykl



dzień – noc jest domeną dla zrozumienia *tydzień*. Analogicznie, trójwymiarowa przestrzeń jest domeną, do której odnosimy rozumienie *w górę i w dół*, podczas gdy trójwymiarowy pojemnik jest domeną dla *na zewnątrz i wewnątrz*.

W zasadzie jakakolwiek konceptualizacja czy konfiguracja wiedzy, nieistotne czy prosta czy złożona, może służyć jako kognitywna domena do opisu znaczeń. Również kategorie morfologiczne i syntaktyczne mogą być interpretowane w odniesieniu do stosownych domen. Na przykład, zdrobnienie, w swoim podstawowym sensie, presuponuje domenę rozmiarów fizycznych, domeną dla zrozumienia gramatycznego czasu przeszłego (lub odniesienia do czasu, który minął) jest czas itd.

W interpretacji kognitywnej idealne modele konceptualne odnoszą się do całej wiedzy człowieka o świecie i wyodrębniane są na podstawie określonych modeli kulturowych, światopoglądowych, poznawczych, stereotypów związanych z przesądami czy uprzedzeniami itp. Przykład jednostki leksykalnej charakteryzowanej przez odniesienie równocześnie do kilku domen podaje Lakoff (jakkolwiek nie stosuje on terminu *domena*), omawiając słowo *matka* (ang. *mother*). Lakoff wskazuje, że aby wyczerpująco zrozumieć to słowo, należy wziąć pod uwagę co najmniej pięć domen. Są to:

- domena genetyczna: matka jest istotą żeńską wnoszącą materiał genetyczny dla dziecka;
- domena narodzin: matka jest istotą żeńską, która rodzi dziecko;
- domena żywienia: matka jest dorosłą istotą żeńską, która karmi i wychowuje dziecko;
- domena genealogiczna: matka jest najbliższym przodkiem rodzaju żeńskiego;
- domena małżeństwa: matka jest żoną ojca.

Możemy porównać domeny, przez odniesienie do których rozumiemy słowo *matka* z domenami niezbędnymi do pełnego zrozumienia słowa *ojciec* (ang. *father*). Wydaje się, że tu znowu zaangażowanych jest pięć domen:

- domena genetyczna: ojciec jest istotą męską wnoszącą materiał genetyczny dla dziecka;
- domena odpowiedzialności: ojciec jest odpowiedzialny finansowo (materialnie) za byt matki i dziecka;
- domena autorytetu: ojciec jest osobą z autorytetem, odpowiedzialną za dyscyplinę dziecka;
- domena genealogiczna: ojciec jest najbliższym przodkiem rodzaju męskiego;
- domena małżeństwa: ojciec jest mężem matki.

Z zestawienia tego wynika, że znaczenie *ojciec* nie jest, z wyłączeniem cech [MĘSKI]/[ŻEŃSKI], identyczne ze znaczeniem *matka*, jak to utrzymywaliby strukturaliści. Obydwa pojęcia są równoważne jedynie w domenach genetycznej, genealogicznej i małżeńskiej.

Idealny model konceptualny znaczenia realizuje cechy ze wszystkich domen, inne modele mogą być niepełne (*matka pracująca*) lub rozszerzone (*dobra matka, choć nie umie prowadzić domu*). Idealne modele kognitywne bywają uściślane cechami stereotypowymi. Cechy ujęte w kognitywnym modelu znaczeniowym mogą być źródłem metafor – *komisja-matka*, czyli „proponująca (rodząca) kandydatów w wyborach”; zaś od cech stereotypowych powstają metonimie odnoszące się do funkcji *matka* – „najważniejszy gracz (najważniejsza osoba) w drużynie palanta”<sup>31</sup>.

Kognitywiści uznają metonimizację i metaforyzację za główną treść języka, co w konsekwencji prowadzi do odrzucenia podziału na wiedzę encyklopedyczną (pozajęzykową) i językową, ponieważ – jak twierdzą – człowiek posługuje się językiem w kontekście różnych, nierozdzielnych rodzajów wiedzy. Tymczasem językoznawstwo autonomiczne zakłada rozdział między wiedzą mówiącego o rzeczywistości a jego wiedzą czysto językową. Znaczenie uważane jest za, w pewnym sensie, niezależne od tego, jaki jest stan rzeczy w świecie rzeczywistym lub wyobrażanym, a także niezależne od tego, co użytkownik języka może wiedzieć o stanach rzeczy.

Z tym ujęciem wiąże się szereg problemów, wśród których rozgraniczenie wiedzy nie jest problemem jedynym. Innym jest pytanie o to, gdzie i na podstawie jakich kryteriów można zakreślić granicę między tym, co jest mówiącemu znane na mocy jego znajomości języka i tym, co wie na podstawie swojej znajomości świata. Problem ten przejawia się szczególnie wyraziście w związku z rozróżnianiem wiedzy syntetycznej i analitycznej. Posługując się tym rozróżnieniem, przypuszczalnie można powiedzieć, że zdanie *Psy są zwierzętami* jest zdaniem analitycznym, czyli że zdanie jest prawdziwe na mocy koniunkcji cech semantycznych<sup>32</sup> definiujących (czysto językowe) znaczenia słów *pies* i *zwierzę*. Jeśli tak, to czy zdania typu *Psy mają cztery nogi. Psy szczekają. Psy nie miauczą. Psy nie mają rogów* są prawdziwe na mocy znaczenia *pies*, czy też na mocy wiedzy o tym, jakie zwykle są psy? Taylor<sup>33</sup> utrzymuje, że aby stwierdzić, że powyższe zdania są analityczne, konieczne byłoby dodanie cech [POSIADAĆ CZTERY NOGI], [SZCZEKAĆ], [-MIAUCZEĆ], [-POSIADAĆ ROGI] itd., tak żeby ustalić warunki niezbędne i wystarczające do zdefiniowania kategorii *pies*. Lecz dodanie tych cech oznacza otwarcie drogi dla nieskończonej długiej definicji tej kategorii; definicja musiałaby zawierać cechy [MIEĆ SIERŚĆ], [MIEĆ PYSK], [-MIEĆ SKRZYDŁA] i wiele, wiele innych. Być może pewnym

<sup>31</sup> Wszystkie przykłady w tym akapicie zostały wzięte z: Encyklopedia 1993a, s. 310.

<sup>32</sup> Termin cecha użyty jest by odwołać się do warunków przynależności do kategorii klasycznych; w odniesieniu do warunków przynależności do kategorii kognitywnych staramy się używać terminu atrybut.

<sup>33</sup> Taylor 1991, s. 81-83.

rozwiązaniem byłoby ograniczenie czysto językowej, słownikowej definicji *pies* do „zwierzę psiego rodzaju”, a więc do zbioru cech [ZWIERZĘ] i [PSI], i pozostawienie zoologom całej reszty wiedzy dotyczącej psów. Sensowne wydaje się bowiem stwierdzenie, że posiadanie czterech nóg jest zaledwie spodziewanym atrybutem psów, nie zaś cechą definiującą kategorię (pies, któremu amputowano nogi, pozostaje psem). Taylor tak pisze: „Nasze przeświadczenie, że zdanie (1) [*Psy są zwierzętami* – przypis JW] jest koniecznie prawdziwe, jest zaledwie konsekwencją naszego głęboko zakorzenionego przekonania o wewnętrznej konstrukcji naturalnych rodzajów, jak pies i zwierzę, oraz o sposobie, w jaki rodzaje naturalne są powiązane w układzie taksonomicznym. Gdyby zaszło zdarzenie wysoce nieprawdopodobne, lecz niewykluczone w sposób absolutny, to jest odkrycie psów jako samoreprodukujących się automatów sterowanych przez pozaziemską myśl, zdanie (1) przestałoby być prawdziwe. Zatem (1) nie może być analityczne, poprzez to, że jego prawdziwość nie jest niezależna od naszego pojmowania nieprzewidywanych (ewentualnych) zdarzeń w rzeczywistym świecie”<sup>34</sup>.

Rozważania te prowadzą w sposób nieunikniony do ujmowania znaczenia słów jako, w szerokim zakresie, encyklopedycznego. Rozumienie *psa* nie jest niezależne od wiedzy o psach – o statusie psów jako gatunku w królestwie zwierząt, o rozmaitych rasach psów, o ich wyglądzie i zachowaniu, ich związku z człowiekiem itd. Ten rodzaj wiedzy jest matrycą dla mentalnej reprezentacji *psa*, a także dostarcza rozumowego uzasadnienia do ograniczenia szeregu stworzeń, które można uznać za psy. Podejście encyklopedyczne również niesie ze sobą problem rozgraniczenia. Prawdopodobnie nie wszystko, co dany człowiek wie o psach, będzie przydatne w charakteryzowaniu jego pojęcia „*psa*”. Stwierdzenie, że słownik ma encyklopedyczny charakter nie jest równoznaczne ze stwierdzeniem, że słownik jest encyklopedią. Równocześnie, nie powinno się a priori wykluczać możliwości posiadania przez różnych użytkowników języka nieco odmiennych reprezentacji „*psa*”. Do indywidualnych charakterystyk „*psa*” równie dobrze może należeć specjalistyczna wiedza hodowcy psów czy weterynarza, jak i wielostronna wiedza miłośnika psów (jak również awersja silnie odczuwana przez kogoś, kto nie lubi psów). Jednak ogólnie biorąc, relewantną informację stanowiącą tło charakterystyki znaczeń słów można traktować jako sieć wspólnej, skonwencjonalizowanej, być może w pewnym stopniu wyidealizowanej wiedzy, wtopionej we wzorce uwarunkowanych kulturowo przekonań i zachowań.

Dostrzeżenie w świecie wyłącznie różnic mogłoby doprowadzić do chaosu percepcyjnego, podczas gdy traktowanie pewnych obiektów jako równoważnych, tj. przynależnych do pewnej, niekiedy specyficznie pojmowanej, klasy równoważności, a zarazem różnych od innych czyni pracę aparatu poz-

---

<sup>34</sup> Taylor 1991, s. 82.

nawczego efektywniejszą. Zdolność ta określa istotę zjawiska i umiejętności kategoryzacji. Wyodrębniane są kategorie tworzące hierarchie zależności, a zasady ich wyodrębniania i hierarchizacji są ważną domeną badań filozoficznych, psychologicznych, antropologicznych, językoznawczych i innych.

Dla psychologii szczególnie interesujące są tzw. pojęcia naturalne powstające w wyniku kategoryzacji codziennego doświadczenia człowieka<sup>35</sup>. Jak pisze Ida Kurcz: „Badania nad tworzeniem się takich pojęć rozpoczęły się w psychologii stosunkowo późno, tj. dopiero w latach siedemdziesiątych. Przez wiele lat psychologowie interesowali się przede wszystkim tworzeniem się pojęć typu zbioru zupełnego o cechach koniecznych i wystarczających, jak np. pojęcie kwadratu. Pojęcia naturalne, kategoryzujące codzienne doświadczenie ludzkie, a więc świat zwierząt i roślin, jak też wytwory działalności człowieka (meble, ubrania itp.), nie dają się już tak łatwo zdefiniować (np. jakie są cechy konieczne i wystarczające owocu?). Granice takich pojęć są płynne, nieostre (np. czy pomidor jest owocem czy jarzyną?), jak też występuje pewnego rodzaju gradacja wewnątrz kategorii pojęciowej (np. jabłko wyda się nam lepszym egzemplarzem owocu niż granat lub kiwi), co wydaje się nonsensem logicznym (coś albo jest owocem, albo nim nie jest; tak, jak dana figura albo jest kwadratem, albo nim nie jest)”<sup>36</sup>.

Liczne przykłady w językach naturalnych pokazują, że znaczeniowa struktura znaku językowego, jakkolwiek by pojęcie znaczenia było rozumiane, może być ujmowana w kategoriach podobieństw i różnic. Równocześnie z semantycznym utożsamieniem klasy znaków językowych wykazujących podobieństwo dokonuje się wewnętrznego jej zróżnicowania poprzez przypisanie każdemu z elementów innego szczegółowego zbioru cech. Dzięki temu znak językowy staje się rozpoznawalny na tle innych znaków zbioru i łączony z klasą obiektów rzeczywistości pozajęzykowej, które oznacza. Poszukiwanie podobieństw i różnic oraz wiążących się z nimi „podobieństw wśród różnic”,

---

<sup>35</sup> Dla potocznej (nawnej) kategoryzacji świata charakterystyczna jest kolektywność oraz typowość. Kolektywność, czyli łączenie podobnych elementów w zespoły „naturalne” a nie „logiczne”, zwane przez psychologów kompleksami lub kolekcjami, jest podstawą funkcjonowania charakterystycznych dla języka potocznego tzw. pojęć kolektywnych, takich jak meble, owoce, odzież itp. Kolekcje, podobnie jak kategorie prototypowe, są odbiciem obrazu świata w umyśle człowieka, czyli kategoriami potocznej wiedzy o świecie w przeciwieństwie do kategorii klasycznych uznawanych za kategorie nauki. Kolekcje mogą mieć zróżnicowaną strukturę wewnętrzną, odpowiednio do ważności elementów dla posługującego się językiem. Myślenie w kategoriach typu to taki sposób porządkowania świata, który polega na wydobywaniu z zespołu przedmiotów jednego najbardziej wyrazistego, zaczynającego następnie funkcjonować jako najlepszy reprezentant całego zespołu, jego prototyp, a dalej na definiowaniu prototypu przez zestaw atrybutów utrwalonych, choć niekoniecznie dystynktywnych. Na przykład zapach róży, kruchość filiżanki, szczekanie psa są to cechy przysługujące nie wszystkim różom, filiżankom, psom, nie są więc kategoriale w sensie logicznym, są jednak cechami typowymi róż, filiżanek, psów, i zarazem stereotypowymi, tj. utrwalonymi w społecznej świadomości.

<sup>36</sup> Kurcz 1995, s. 225-226.

czyli uniwersaliów, jest wspólne wszystkim koncepcjom językoznawczym, choć różne koncepcje mogą je lokować w różnych sferach (języka, bytów idealnych, umiejętności percepcji świata i przetwarzania informacji)<sup>37</sup>.

Podobieństwo jako podstawowa relacja organizująca znaczenie przejawia się nie tylko w zaliczaniu całej grupy znaków językowych do nadrzędnej klasy semantycznej. Również sugeruje możliwość przyjęcia dla całej grupy wspólnego modelu definicyjnego. Ten kierunek zaistniał w światowej literaturze językoznawczej dzięki Annie Wierzbickiej. Wierzbicka przyjęła założenie, iż słownictwo należące do tej samej grupy leksykalnej nie tylko realizuje podobny model definicji semantycznej, ale także hierarchia poszczególnych składników definicyjnych nie jest przypadkowa, lecz podporządkowana sposobowi myślenia o rzeczywistości. Gdy człowiek styka się z jakimś świadomym wytworem rąk ludzkich (artefaktem), to przede wszystkim zadaje pytanie: do czego ten przedmiot służy, jakie jest jego przeznaczenie? Dopiero określenie funkcji pozwala ocenić właściwości fizyczne, kształt, wielkość, barwę itp. tego obiektu. Funkcja jako hierarchicznie najważniejszy składnik semantyczny definicji artefaktów winna być kluczem do poszukiwania dalszych składników jego znaczenia. Prace Wierzbickiej mają duże znaczenie nie tylko dla praktyki opisów znaczeniowych. Stanowią one ponadto przykład podejścia do struktury znaczeniowej języka, w której decydująca o przynależności do tej samej grupy leksykalno-semantycznej kategoria podobieństwa widoczna jest nie tylko we wspólnym elemencie *genus proximum*<sup>38</sup>, lecz również we wspólnym lub podobnym modelu definicyjnym tej grupy.

W procesie kategoryzacji sprawą podstawową jest jednak nie tylko model definicyjny, ale także sposób ujmowania cech istotnych w znaczeniu słowa, inaczej mówiąc, związek między sposobem myślenia o świecie a zestawem cech konkretnego wyrazu w konkretnym układzie kategoryalnym. „Jest oczywiście sprawą niezwykle ciekawą jak to się dzieje, że z ogromnego układu treści rzeczy <<bierzemy>> bezwolnie w pierwszym etapie poznawczym te właśnie cechy-elementy, a nie inne. Można w tej materii snuć różne hipotezy, a wśród nich i teoria o naszym układzie genowym będącym rezultatem rozwo-

---

<sup>37</sup> Współcześni kognywiści nie odrzucają klasycznej arystotelesowskiej zasady dokonywania porównań: szukania podobieństw i różnic między rzeczami. Dla nich także porównywanie jest podstawą wszelkiej kategoryzacji. Podkreślają natomiast, że procesy poznawcze są ograniczane przez wiele niejednorodnych czynników (indywidualne cechy osobowości, jednostkowe zdolności, poziom wykształcenia, system przekonań, motywacji i wartości, status społeczny, kontekst kulturowy itp.) warunkujących percepcję. Tak więc percepcja jest z konieczności procesem wybiórczym: z jednej strony ograniczają ją możliwości naszych zmysłów, z drugiej – nastawienie psychiczne.

<sup>38</sup> Klasyczna zasada definiowania poprzez *genus proximum* i *differentia specifica* nie jest wprawdzie zasadą uniwersalną, którą można by stosować do wszystkich bez wyjątku znaków języka, ale może być stosowana do wiclu, jeśli nie większości, wyrazów. *Genus proximum* łączy przez wyrażenie o charakterze nadrzędnym elementy w istocie różne, ale wykazujące jakąś cechę wspólną. Cechy szczegółowe (*differentia specifica*), podtrzymując to podobieństwo, jednocześnie wewnątrznie różnicują analizowaną grupę znaków językowych.

ju w historii, nie jest hipotezą do pogardzenia. Następnie uwzględnić trzeba także momenty kultury (...) Zwrócić wszakże należy uwagę także i na to, że nie jesteśmy tylko tworem natury i historycznym produktem. Jesteśmy bytem osobowym, w którym, jak to rejestrujemy w naszym doświadczeniu siebie, zachodzą niezdeteminowane, wolne akty poznania wybiórczego i miłości. (...) Na akty naszego poznania czysto teoretycznego ogromny wpływ ma nadto nastawienie psychiczne (...), które w znacznej mierze decyduje o doborze poznanych treści. Dwaj ludzie oglądający ten sam obiekt z różnymi psychicznymi uzdolnieniami i zamiłowaniem widzą swoiście, ujmując rzeczy cechy różne, lub w różnej relacji do innych komponentów treściowych<sup>39</sup>. Krapiec jest też autorem następującego znaczącego stwierdzenia: „Świat otaczający człowieka nie był traktowany jako przedmiot poznania teoretycznego – by wiedzieć dla samej wiedzy – ale jako przedmiot poznania praktycznego – aby być użytym jako narzędzie służące do przetrwania<sup>40</sup>. Jeśli przyjąć, że istotnie poznanie ma najczęściej celowościowy, a nie stricte poznawczy charakter (co pokrywałoby się z pewnymi hipotezami dotyczącymi istoty motywacji ludzkich zachowań<sup>41</sup>), to należałoby dopuścić możliwość istnienia w danym momencie czasu wielu „prawdziwych” obrazów porządku świata, zarówno naukowych, jak i potocznych. W literaturze z zakresu nauki o informacji wątek ten pojawił się już jakiś czas temu, choć trzeba przyznać, że był w większym stopniu oparty na intuicji i wybiórczej obserwacji niż na dowodzie naukowym czy choćby wyjaśnieniu uprawdopodobniającym. Jego konsekwencją dla teorii i praktyki języków informacyjno-wyszukiwawczych powinno być zróżnicowanie jiw, stosowne do owej wielości „prawdziwych” obrazów świata.

Obrazy świata tworzone przez język nauki i język ogólny są często zbudowane odmiennie. „Obraz naukowy kształtowany jest przede wszystkim przez odniesienia do rzeczywistości, a przynajmniej dąży do możliwie wiernego jej odwzorowania. Obraz świata zawarty w języku ogólnym, zwany niekiedy obrazem naiwnym lub potocznym, uwzględnia zarówno obiektywnie dane cechy rzeczywistości, jak i kulturowe uwarunkowania rządzące myśleniem i ocenami człowieka. Naukowe poznanie świata i pochodny od niego naukowy obraz są zasadniczo wspólne ludziom mówiącym różnymi językami, z kolei potoczny obraz świata jest utrwalony i odtwarzany poprzez język ogólny i dlatego każdy język narodowy ma sobie tylko właściwy obraz świata<sup>42</sup>. Dodajmy, że język naturalny jest w praktyce tworzywem do wyrażania wielu obrazów świata wyróżnionych ze względu na różne kryteria psychologiczne, socjologiczne, dziedzinowe itd. „Pojęcia naukowe są definiowane wyłącznie

<sup>39</sup> Krapiec 1985, s. 183-184.

<sup>40</sup> Ibidem s. 185.

<sup>41</sup> Tokarz 1985.

<sup>42</sup> Tokarski 1993, s. 339.

w kategoriach metafizycznych<sup>43</sup> (por. np. definicję kategorii KWARK). Wiele kategorii dopuszcza oba rodzaje definicji (por. np. naukową i potoczną definicję PRÓŻNI). Inne wreszcie definiowalne są wyłącznie w terminach wiedzy potocznej (por. np. definicję kategorii DOM)<sup>44</sup>.

Akty kategoryzacji, zwłaszcza jeśli dotyczą potocznego obrazu świata, nie mogą nie uwzględniać elementu wartościowania (oceny) jako immanentnego składnika znaczenia znaku językowego. Inaczej mówiąc, kategoryzacja potoczna angażuje zarówno kryteria poznawcze, obiektywne, jak i emocjonalne, oparte na wartościowaniu. Wyrażanie ocen w języku może się dokonywać albo bezpośrednio za pomocą wyrazów nazywających negatywne lub pozytywne oceny i ich aspekty, albo pośrednio, niejawnie, co wymaga dodatkowych działań interpretacyjnych. W nauce polskiej próby systematycznych analiz wartości pojawiły się najpierw w filozofii i socjologii, zwłaszcza w pracach Marii Ossowskiej i Władysława Tatarkiewicza, znacznie później w językoznawstwie, w którym problematyką tą zajęła się przede wszystkim Jadwiga Puzynina<sup>45</sup>. Szczególnego znaczenia nabrało zagadnienie wartościowania w językoznawstwie kognitywnym. „Definicje kognitywne nazw rodzajów i gatunków naturalnych opierają się na wnikliwych studiach nad językiem naturalnym oraz sposobami, jakimi ludzie kategoryzują i konceptualizują rzeczywistość w poznaniu, co znajduje swoje potwierdzenie w używaniu danej nazwy. (...) [Na przykład] Podejmując się zadania prezentacji znaczenia pojęcia PIES, nie sposób zapomnieć, że w wielu wypadkach mówimy o czymś więcej niż o psie, por. konstrukcje *psi, pieski*, np. *psi żywot, pieskie życie*, gdyż znaczenie „należący do psa, właściwy psu, (SJPD<sup>46</sup> [przypis – JW] VI: 22) wprowadza nas nie tylko w świat etnometodologii i etnobiologii, lecz także w świat wartościowanych negatywnie ludzkich ocen zjawisk i osób”<sup>47</sup>. Wartościowanie językowe jak do tej pory nie było przedmiotem wielu publikacji z dziedziny języków informacyjno-wyszukiwawczych. Tymczasem doświadczenie pokazuje, że w językach tych, jak w soczewce, mogą się skupiać wartości, stereotypy, uprzedzenia, sympatie i antypatie różnych grup społecznych oraz że jiw, zwłaszcza te o paranaturalnej notacji, mogą łatwo stać się narzędziem propagandy i indoktrynacji<sup>48</sup>.

W klasycznych rozważaniach poświęconych kategoryzacji wielu językoznawców przyjmuje podejście wykazujące analogie z założeniami, terminologią, a nawet z notacją<sup>49</sup> stosowaną w fonologii. Mamy tu do czynienia z tym, co bywa nazywane założeniem strukturalnej analogii: „Jest to po prostu

---

<sup>43</sup> W tym kontekście *kategoria metafizyczna* znaczy ‘kategoria klasyczna’, ‘realistyczna’, ‘odpowiadająca rzeczywistości światu’. [Przypis – JW].

<sup>44</sup> Tabakowska 1995, s. 46.

<sup>45</sup> Tokarski 1993, s. 344.

<sup>46</sup> SJPD, s. 22.

<sup>47</sup> Nowakowska-Kempna 1996, s. 53-59.

<sup>48</sup> Przykładów potwierdzających tę tezę dostarcza m.in. praca: Miller 1980.

<sup>49</sup> Z fonologii pochodzi konwencja notacyjna polegająca na zapisie cech dystynktywnych i atrybutów kapitalikami i ujmowaniu ich w nawiasy kwadratowe.

założenie znane z wielu prac postsaussure'owskich, iż należy oczekiwać powtarzalnego występowania (rekurencji) tych samych cech na różnych poziomach. Własności strukturalne przyjmowane za unikatowe dla określonego poziomu są niespodziewane i wręcz podejrzone, o ile nie są wsparte mocnym dowodem na ich unikatową odpowiedniość w tym konkretnym wypadku”<sup>50</sup>. Skoro zatem kategorie fonologiczne można reprezentować za pomocą cech binarnych, elementarnych, uniwersalnych, abstrakcyjnych i wrodzonych, to można się spodziewać, że i kategorie semantyczne można charakteryzować podobnie, za pomocą cech binarnych, elementarnych, uniwersalnych, abstrakcyjnych i wrodzonych.

Analiza kategorii semantycznych według zasad znanych z fonologii była uprawiana w ramach paradygmatu transformacyjno-generatywnego, a także pod nazwą analizy składnikowej (komponencjalnej) przez nie-generatywistów. „Dziś już właściwie nie ma badań semantycznych nie przyjmujących tego postulatu metodologicznego, a więc można powiedzieć, że analiza składnikowa (komponencjalna) stanowi podstawę wszelkiego opisu semantycznego”<sup>51</sup>. Pierre Guiraud tak objaśnia istotę analizy składnikowej: „celem tej nowej semantyki jest rekonstrukcja <<systemu sensów>> na podobieństwo <<systemu dźwięków>> zbudowanego przez fonologię. Skoro bowiem wyraz w planie elementów znaczących jest <<pękiem fonemów>>, przeto można by sobie wystawić, że również pojęcie znaczone jest <<pękiem elementów jednostek znaczenia>>, i tak krowa byłaby tworzona przez zwierzę + samica + domowe + itd. Stąd nazwa <<analiza składnikowa>> (czyli rozbiór na składniki semiczne), nadawana tego typu badaniom”<sup>52</sup>. Jedną z pierwszych prac ukazujących opis znaczenia wyrazu metodą wyróżniania jego elementów znaczących, zwanych przez niektórych językoznawców semami lub sememami, było studium Bernarda Pottiera<sup>53</sup> poświęcone analizie znaczeniowej wyrazów nazywających meble służące do siedzenia. Przytoczymy ważniejsze elementy tej kategoryzacji, stwarzając w ten sposób tło porównawcze dla cytowanej dalej kategoryzacji prototypowej mebli.

W zbiorze wyrazów nazywających meble do siedzenia Pottier wyróżnił sześć składników (cech-semów), których obecność lub brak spowodowały wyodrębnienie się nazw pięciu typów mebli – *krzesło*, *fotel*, *taboret*, *kanapa* i *puf*. Były to następujące cechy-semy ( $S_i$ ):

- $S_1$  = ‘z oparciem’
- $S_2$  = ‘na nogach’
- $S_3$  = ‘dla jednej osoby’
- $S_4$  = ‘do siedzenia’
- $S_5$  = ‘z poręczami’
- $S_6$  = ‘szytwna konstrukcja’

<sup>50</sup> Anderson 1986, s. 3. Cyt. za: Taylor 1991, s. 29.

<sup>51</sup> Grzegorzczkova 1995, s. 78.

<sup>52</sup> Guiraud 1976, s. 100. Cyt. za: Bartol-Jarosińska 1996, s. 316.

<sup>53</sup> Pottier 1963. Cyt. za: Bartol-Jarosińska 1996, s. 317-318.



| Mebel   | S <sub>1</sub> | S <sub>2</sub> | S <sub>3</sub> | S <sub>4</sub> | S <sub>5</sub> | S <sub>6</sub> |
|---------|----------------|----------------|----------------|----------------|----------------|----------------|
| krzesło | +              | +              | +              | +              | -              | +              |
| fotel   | +              | +              | +              | +              | +              | +              |
| taboret | -              | +              | +              | +              | -              | +              |
| kanapa  | +              | +              | -              | +              | +              | +              |
| puf     | -              | +              | +              | +              | -              | -              |

Znaczenie każdego z wyodrębnionych wyrazów można określić w stosunku do znaczenia innych wyrazów przez wskazanie obecności lub braku którejsz z sześciu cech. W ten sposób *krzesło* = ‘z oparciem’ + ‘na nogach’ + ‘dla jednej osoby’ + ‘do siedzenia’ + ‘szywna konstrukcja’. Cecha ‘z oparciem’ odróżnia znaczeniowo *krzesło* od *taboret* i *puf*; cecha ‘szywna konstrukcja’ odróżnia od *pufa* wszystkie pozostałe rodzaje mebli itd. Z tabeli wynika, że cechy ‘na nogach’ i ‘do siedzenia’ stanowią cechy wspólne wszystkich pięciu nazw mebli, pozostałe natomiast są cechami specyfikującymi (dystynktywnymi), które odróżniają od siebie znaczenia poszczególnych wyrazów, należących do przedstawionego zbioru.

Podjęcie do kategorii semantycznych oparte na cechach dystynktywnych odwołuje się do argumentów tego samego rodzaju, jakie stosowane są w fonologii. Według Taylora<sup>54</sup> zalety tego podejścia można widzieć trojako:

1) umożliwia ustalenie zachodzących w słowniku pewnych związków relacyjnych. Na przykład słowa *kawaler* i *panna* są niewątpliwie powiązane, podobnie jak związane są, choć w inny sposób, fonemy /b/ i /p/. Co więcej, relacja między *kawaler* i *panna* w jakiś sposób odpowiada związkom zachodzącym między innymi wyrazami, takimi jak na przykład *chłopak* i *dziewczyna*, *mąż* i *żona*, *wuj* i *ciotka* itp. Wyrazy w poszczególnych parach różnią się między innymi tym, że jeden ma cechę [MĘSKI], a drugi – [ŻEŃSKI]. Poza tym specyfikacja cech dla obu wyrazów pary jest identyczna. Analiza oparta na cechach dystynktywnych umożliwia uchwycenie różnych relacji zachodzących między wyrazami, np. relacji inkluzji czy relacji hiponimii. Znaczenie słowa *mężczyzna* z cechami [ISTOTA LUDZKA], [DOROSŁY], [MĘSKI] zawiera się w znaczeniu *kawaler*. *Mężczyzna* jest zatem nadrzędny w stosunku do *kawaler*, a *kawaler* jest hiponimem *mężczyzna*;

2) cechy dystynktywne umożliwiają definiowanie naturalnych klas obiektów. I tak, cecha [ISTOTA LUDZKA] definiuje klasę nazw istot ludzkich, podczas gdy [NIEOŻYWIONY] wyznacza klasę nazw obiektów nieożywionych. Tego rodzaju klasy związane są z określaniem ograniczeń dotyczących łączenia wyrazów w struktury wyższego rzędu. Na przykład, podmiotem lub dopełnieniem czasownika nie może być dowolny rzeczownik – specyfikacja semantyczna czasownika wymaga określenia klasy rzeczowników, które mogą funkcjonować jako jego podmiot czy dopełnienie. Nie można powiedzieć *Dobroć podziwia Jana*, ponieważ *podziwiać* wymaga jako podmiotu rzeczono-

<sup>54</sup> Taylor 1991, s. 29-37.

wnika z cechą semantyczną [ISTOTA LUDZKA]. Podobnie, w połączeniach przymiotnik-rzeczownik konieczna jest kongruencja (związek zgodny) między specyfikacjami cech dystynktywnych wyrazów składowych konstrukcji. Nie można mówić o *niewowleńczym kawalerze* z cechą [-DOROSŁY], gdyż pozostaje w sprzeczności z *kawaler*, które to słowo zawiera cechę [DOROSŁY];

3) semantyczne cechy dystynktywne wyjaśniają pewne typy znaczeniowe zdań oraz związki zachodzące między zdaniem, na przykład relacje implikacji, synonimii, sprzeczności. Za pomocą cech dystynktywnych można wyjaśnić i inne aspekty znaczenia zdań, na przykład, współgranie specyfikacji cech może pomóc w wyjaśnianiu skutków łączenia zdań pewnymi spójnikami.

Według Taylora analiza składnikowa nie ma jeszcze możliwości opisu słownictwa języka naturalnego w szerszym wymiarze. Wykorzystywana jest najczęściej w pracach językoznawczych o charakterze teoretycznym i dotyczy wybranych wyrazów. Dane, którymi posługują się semantycy uprawiający analizę składnikową, rzadko wykraczają poza poziom złożoności reprezentowany przez przykłady typu *kawaler* czy *meble służące do siedzenia*.

Z analogią w stosunku do fonologii posunięto się dalej. Tak, jak w fonologii traktuje się cechy dystynktywne fonemów jako jednostki elementarne, tak też i w semantyce niektórzy badacze twierdzą, że cechy semantyczne są ostatecznymi, atomowymi jednostkami składającymi się na znaczenie słów. Teoria semantycznych cech dystynktywnych, wywodząca się od Hjelmsleva, stała się podstawą wielu koncepcji w językoznawstwie. Wzbogaconym źródłem koncepcji komponentalnej były tezy semantyki generatywnej, która wprowadziła do opisu semantycznego nowy aparat – logikę symboliczną i przede wszystkim rachunek predykatów. Istotą tych poszukiwań było przekonanie, że znaczenie nie stanowi sumy składników (paralelnych do fonologii), ale zhierarchizowane struktury, posiadające wewnętrzną składnię.

Składnikowa koncepcja znaczeń istotnie wzbogaciła współczesne badania językoznawcze. Dzięki niej teoria pól semantycznych, która w okresie bezpośrednio powojennym została nieco usunięta w cień, zaczęła przeżywać swój renesans. U podstaw teorii pól znaczeniowych<sup>55</sup> leży systemowe widzenie leksyki danego języka, tj. odrzucenie zatowizowanego sposobu widzenia słowa, jego znaczenia i zmian zachodzących w rozwoju języka. Ważnym teoretykiem pól semantycznych był Leo Weisgerber, który w swoich pracach nawiązywał do idei „wewnętrznej formy języka” Wilhelma von Humboldta<sup>56</sup>, organizującej widzenie świata przez daną społeczność językową. Wewnętrzna forma narzuca ludziom mówiącym danym językiem określony sposób

---

<sup>55</sup> Uporządkowany przegląd zagadnień wiążących się z metodologią pól znaczeniowych przedstawiają między innymi następujące prace: Buttler 1967; Grzegorzczakowa 1995; Lyons 1984-1989; Miodunka 1980; Pisarczuk 1967.

<sup>56</sup> Humboldtowską tezę o narzucaniu przez język widzenia świata potwierdziły późniejsze badania etnolingwistyczne, psychologiczne i antropologiczne, choć podstawowe tezy neohumboldtysty pozostają ciągle w sferze hipotez.

widzenia rzeczywistości. Pole wyrazowe jest odbiciem pola semantycznego, które stanowi swoistą siatkę pojęciową, nakładaną przez język na rzeczywistość. Owe siatki pojęciowe są różne w różnych językach, a także zmieniają się w różnych okresach rozwoju tego samego języka.

Słownictwo języka stanowi zatem zwarty system wzajemnie ze sobą powiązanych leksemów. Ten całościowy system rozpada się na podsystemy, zwane polami znaczeniowymi (semantycznymi, pojęciowymi), w których znaki językowe i ich podzbiory wzajemnie wyznaczają swoje granice znaczeniowe. Każda zmiana w obrębie pola, np. usunięcie jakiegoś elementu czy wprowadzenie nowego, musi spowodować semantyczne przekształcenia elementów pozostałych. Kształt poszczególnych pól znaczeniowych, tj. ich zakres i wewnętrzne relacje między jednostkami, jest wyrazem rozumienia i widzenia świata właściwego dla danej grupy użytkowników języka. W konsekwencji prowadzi to do przekonania o istnieniu abstrakcyjnej siatki znaczeń nakładanej na rzeczywistość.

Praktyka badań pól językowych skupiała się zwykle na szczegółowych analizach wybranych fragmentów a nie na całych systemach leksykalnych. Dają się tu zaobserwować dwie tendencje zbieżne z rozróżnieniem związków syntagmatycznych i paradygmatycznych. Na analizach związków semantycznych grup słów, które są ze sobą powiązane składniowo, np. *chwytać – ręka*, *szczekać – pies*, *blond – włosy* itp. koncentrowała się, wywodząca od Porziga, metodologia badań pól syntagmatycznych. Natomiast analiza związków paradygmatycznych, wykorzystująca zasadę współprzynależności słów do tej samej klasy znaczeniowej na podstawie łączącego je semantycznego podobieństwa, a wywodzona od badań Josta Triera, obejmowała grupy leksykalne składające się z jednostek należących do jednej kategorii części mowy (np. nazwy określonych typów czynności, nazwy ludzi, ptaków itd.). Nurt paradygmatyczny pokazywał wzajemną zależność jednostek należących do tego samego pola, dynamikę przekształceń całego pola w efekcie zmiany choćby jednego z elementów składowych i, wreszcie, zależność budowy pola odmiennych językowych obrazów świata różnych języków.

Istotnymi problemami badań pól językowych są granice i skład leksykalny pola oraz kryteria językowego uporządkowania zbioru wyrażań w sposób odpowiadający postrzeganiu świata przez użytkowników języka. Zagadnienie granic i składu leksykalnego wiąże się po części z arbitralną i subiektywną decyzją badacza, ale w głównej mierze są one wyznaczone przez znaczeniowe aspekty samego języka i językową świadomość jego użytkowników. Arbitralną decyzją opisującego jest wybór fragmentu leksyki będącego zwykle składnikiem całości większej i zarazem składającego się z podzbiorów zakresowo węższych. Hierarchiczna struktura przeważającej części słownika powoduje, że wybór do analizy pewnego pola wymaga jego usytuowania wobec pól nadrzędnych, rozważenia możliwych związków z polami tego samego poziomu oraz wykazania w opisie niższych pięt hierarchii. Wybór

przedmiotu analizy jest zatem zabiegiem wprawdzie koniecznym, ale prawie zawsze nieco sztucznym, gdyż zakreślenie granic pola przerywa naturalne więzi łączące dane pole z innymi.

Znaczeniowe właściwości leksyki powodują też pewne rozmycie granic pola. W obrębie niemal wszystkich analizowanych pól (z wyjątkiem może klas zamkniętych typu 'dni tygodnia', 'pory roku') pewna część słownictwa stanowi centrum tego pola (jest z nim kojarzona najszybciej) i jej odtworzenie nie stanowi poważniejszego problemu. Im dalej od centrum, tym większe problemy z wyznaczeniem precyzyjnej granicy i składu leksykalnego<sup>57</sup>. O płynności granic decyduje również wieloznaczność słowa i jego kontekstowo realizowane cechy semantyczne, tzw. konotacje znaczeniowe. Powodują one, że to samo słowo jest składnikiem przynajmniej dwu różnych pól, a ich rozdzielenie automatycznie gubi więzi łączące różne znaczenie tego samego słowa. Można bowiem czasownik *pełzać* zaliczyć do pola czasowników ruchu, a rzeczownik *osioł* do pola nazw zwierząt, ale poprzez swe znaczenia przenośne słowa te należą również do pól innych, związanych w tym wypadku z kategoriami działań ocenianych negatywnie czy ludzi moralnie lub intelektualnie gorszych. Semantyczna struktura słowa pozwala na płynne przechodzenie z jednego pola do drugiego, a wyizolowanie jednego pola tę płynność przerywa<sup>58</sup>.

Drugim istotnym, poza granicami i składem leksykalnym, definicyjnym elementem pola jest kryterium jego uporządkowania. Układ alfabetyczny nie oddaje ani zależności znaczeniowych, jakie istnieją między jednostkami leksykalnymi, ani swoistości postrzegania świata przez dany język. W obrębie pola można wyróżnić różnego typu relacje znaczeniowe<sup>59</sup>: synonimie, antonimie, komplementarność i konwersję, hiponimie, semantyczną niezgodność, partytywność itd. Wiele z tych kategorii wykorzystywano w opisach języka już wcześniej. Istotnym novum strukturalnego porządkowania pola jest to, że ta sama jednostka leksykalna może występować w polu w wielorakich zależnościach, łącznie wyznaczających jego hierarchię. *Kobieta*, będąc hiponimem nadrzędnej kategorii *człowiek*, jest zarazem elementem komplementarnej pary *mężczyzna – kobieta*, pozostaje w relacji synonimii ze słowem *niewiasta* itd.

Na dwie z tych relacji znaczeniowych, hiponimie i synonimie, warto zwrócić szczególną uwagę, gdyż ich rola w języku jest specyficzna. Hiponimia uwydatnia tę istotną właściwość języka naturalnego, która dąży do włączenia każdej nazwy do klasy semantycznie szerszej, a zarazem otwiera możliwość wydzielenia szczegółowszych podklas jej znaczenia. Wydaje się, że dla przeważającej części słownictwa to porządkujące widzenie elementu językowego przez pryzmat szerszej klasy znaczeniowej jest odbiciem generalnego zjawiska pojęciowego porządkowania świata. O ile hiponimia buduje wertykalne układy

<sup>57</sup> Miodunka 1980, s. 151-155.

<sup>58</sup> Tokarski 1984.

<sup>59</sup> Istotny wkład wniosły tu prace z zakresu semantyki strukturalnej, a zwłaszcza Lyonsa (Lyons 1984-1989).

porządkujące leksykę, o tyle synonimia wprowadza uporządkowanie na poszczególnych poziomach. Warto zwrócić uwagę, że synonimia jest cechą przede wszystkim języka naturalnego, a unikają jej języki sztuczne, np. logiki czy języki informacyjno-wyszukiwawcze. Wprowadza ona głównie zróżnicowanie stylistyczne i ekspresywne, eksponujące emocjonalny stosunek użytkowników języka do wypowiedzanych treści.

Postulowanie uniwersalnych elementarnych cech semantycznych nie jest oczywiście innowacją wniesioną przez strukturalistów czy generatywistów. Już w siedemnastym wieku Leibnitz postawił sobie za zadanie odkrycie „alfabetu myśli ludzkich” – zbioru podstawowych „klocków pojęciowych”, nie podlegających dalszej dekompozycji, których kombinacje mogą leżeć u podstaw wszelkich możliwych pojęć w języku. Śladem tej myśli filozoficznej poszły tzw. słowniki pojęciowe, które zaczęły powstawać w dziewiętnastym wieku. Niektórzy współcześni językoznawcy działający poza paradygmatem generatywnym, łącznie z Anną Wierzbicką i Andrzejem Bogusławskim, przyłączyli się do programu odkrycia owych uniwersalnych cech semantycznych. Jednakże, podczas gdy analogie między dystynktywnymi cechami fonologicznymi i semantycznymi są nieodparte, to istnieje różnica ilościowa i zarazem jakościowa między przypuszczalnymi uniwersalnymi elementarnymi cechami fonologicznymi i domniemanymi „klockami pojęciowymi myśli ludzkiej”. Według Taylora „przy analizie systemu dźwięków jakiegoś języka liczba kategorii fonemów wymagających zidentyfikowania jest dość mała (około 45 w angielskim, zależnie od analizowanego dialektu oraz od zastosowanej metody analizy). Liczba niezbędnych uniwersalnych dystynktywnych cech fonologicznych jest również dość łatwa do ustalenia (Chomsky i Halle sugerują około dwudziestu). W odróżnieniu od fonologii, liczba kategorii semantycznych w dowolnym języku nie tylko jest ogromna, ale także (jak to wykazuje ciągle powstawanie nowych wyrazów) możliwa do powiększania, i stąd też na pierwszy rzut oka wydawałoby się nierealistyczne oczekiwanie umiejętności zredukowania wszystkich możliwych znaczeń słów we wszystkich językach naturalnych do uchwytneho, skończonego zbioru uniwersalnych cech elementarnych”<sup>60</sup>.

Kończąc wyliczanie analogii między fonologicznymi i semantycznymi cechami dystynktywnymi, chcemy jeszcze zasygnalizować, jak dotąd nierozstrzygnięty, problem abstrakcyjności i ewentualnego wrodzonego charakteru cech semantycznych. Wiąże się on z pytaniem o naturę znaczenia znaku językowego i sferę jego interpretacji. Część badaczy zajmujących się tą problematyką opowiada się za hipotezą, że mentalna reprezentacja obiektów i sytuacji świata zewnętrznego nie jest związana ze strukturą semantyczną języka. Podczas gdy pewne mechanizmy rozumowania muszą w oczywisty sposób wiązać znaczenia słowa i zdania z rzeczywistością, to cechy semanty-

---

<sup>60</sup> Taylor 1991, s. 33.

czne charakteryzujące znaczenia słowa i zdania ani nie zastępują jakichkolwiek własności fizycznych i związków, ani nie są możliwe do nauczenia z doświadczenia owych własności i związków. Z tego stwierdzenia wynikałby wrodzony charakter cech semantycznych. Jeśli cechy te są abstrakcyjne, a więc nie jest możliwe ich wyprowadzenie na podstawie wiedzy o rzeczywistości, to aby w ogóle możliwe było przyswajanie języka, konieczne jest założenie, że znajomość zbioru uniwersalnych cech jest dziedziczona genetycznie.

Wielu lingwistów odnosiło się krytycznie do hipotezy o abstrakcyjnym, uniwersalnym, elementarnym i wrodzonym charakterze cech semantycznych. Na przykład u Lyonsa widoczne jest to poprzez określenie znaczenia jako zbioru relacji wewnątrz systemu językowego. Wiedza mówiącego o rzeczywistości pozostaje czynnikiem zewnętrznym wobec systemu językowego. Rozwijając rozróżnienie nakreślone przez Fregego, Lyons wprowadził rozróżnienie między sensem wyrażenia językowego i jego odniesieniem (referencją). Referencja dotyczy desygnowania (oznaczania) obiektów w rzeczywistości. Z kolei sens jest abstrakcyjnym znaczeniem językowym. Dla Lyonsa, sens słowa jest równoważny zbiorowi relacji (synonimii, generycznych, implikacji, niespójności itp.) zachodzących między daną jednostką i innymi jednostkami w systemie leksykalnym. Sens jest niezależny od jakichkolwiek własności rzeczy poza samym systemem językowym. „O sensie mówi się tu jako o związku między wyrazami lub wyrażeniami tego samego języka bez względu na ewentualną relację zachodzącą między nimi a ich odnośnikami czy denotatami. Pytanie: „Jaki jest sens takiego a takiego wyrazu lub wyrażenia?” jest więc bardziej ograniczone, niż pytanie: „Co znaczy dany wyraz lub wyrażenie?”. (...) zarówno o pojedynczych wyrazach słownikowych (...), jak i o wyrażeniach mówi się, że mają sens (i denotację), podczas gdy tylko wyrażenia (i to tylko niektóre z nich) mają odniesienie”<sup>61</sup>. Wynikałoby z tego, że ponieważ sens jest zdefiniowany w kategoriach relacji zachodzących między jednostkami słownika, nie niesie ze sobą żadnych presupozycji o istnieniu obiektów lub własności poza słownikiem rozważanego języka.

Podobne, strukturalistyczne i komponencjalne, podejście do znaczenia leksykalnego (znaczenia słowa) reprezentuje Cruse<sup>62</sup>. Jednakże składniki, które przyjmuje Cruse nie są abstrakcyjnymi elementarnymi cechami semantycznymi. Składnikami znaczenia słowa są po prostu znaczenia innych słów. Cruse uważa, że zamiast mówić, że cecha dystynktywna [ZWIERZE] jest cechą składową znaczenia *pies* wystarczy powiedzieć, że znaczenie słowa *zwierzę* jest „zawarte” w znaczeniu słowa *pies*. Zrezygnowanie z dogmatu abstrakcyjności otwiera drogę do włączenia wiedzy encyklopedycznej do definicji słów. Problem ustalenia granicy między znaczeniem słowa i faktami

<sup>61</sup> Lyons 1984-1989, T. 1, s. 202.

<sup>62</sup> Cruse 1986. Cyt. za: Taylor 1991, s. 35-37.

encyklopedycznymi, czyli między wiedzą językową i wiedzą pozajęzykową był przedmiotem zainteresowania wielu językoznawców. W odniesieniu do języków informacyjno-wyszukiwawczych analizowała go Elżbieta Artowicz<sup>63</sup>.

Kolejna nowość wprowadzona przez Cruse wyraża się w tym, że składniki znaczenia słów nie muszą spełniać warunku bycia niezbędnymi i wystarczającymi. Niektóre składniki rzeczywiście mają charakter kryteriów; bycie zwierzęciem pełni zasadniczą funkcję w znaczeniu *pies*. Inne składniki są wykluczone; np. bycie kotem jest wykluczone ze znaczenia *pies*. Między tymi dwoma biegunami Cruse wskazuje pewne kontinuum wyznaczające sposób, w jaki znaczenie jednego wyrażenia uczestniczy w znaczeniu innego wyrażenia. Tym nie mniej wystąpienia pewnych składników można się spodziewać, choć nie mają kryterialnego charakteru. Na przykład, takim elementem znaczenia *pies* jest zdolność szczekania, co jednak nie znaczy, że niezdolność szczekania jest wystarczającym powodem do wykluczenia kategoryzacji danego obiektu jako psa. Równocześnie inne składniki choć nie wykluczone, nie są spodziewane, np. zdolność śpiewania, która nie wyklucza kategoryzacji obiektu jako psa; raczej możemy mieć do czynienia z niezwykłym psem lub ze zdegradowaną formą śpiewu.

Istnieją rozmaite punkty widzenia, z których można patrzeć na klasyczną teorię kategoryzacji i ewentualnie ją krytykować. Czy wszystkie elementy kategorii koniecznie mają udział w zbiorze cech służących jako kryterium kategoryzacji? Czy znaczenia słów dają się sprowadzić do zbioru elementarnych składników semantycznych? Czy zasadne jest wprowadzanie rozróżnienia między wiedzą językową i pozajęzykową danej osoby? Czy można znać znaczenie słowa niezależnie od czyjejś znajomości relewantnych faktów w rzeczywistości? To tylko kilka pytań, na które trudno udzielić odpowiedzi w ramach tylko jednego (strukturalistycznego) paradygmatu. Problemy te stały się przedmiotem zainteresowania zwolenników różnych podejść do zagadnień języka, między innymi podejścia kognitywnego, którego jednym z wyrazów na gruncie semantyki jest teoria prototypów i kategorii prototypowych. Warto w tym miejscu przypomnieć, że aby znaleźć znaczące uogólnienia, językoznawstwo kognitywne także szuka arystotelesowskich „podobieństw wśród różnic”, czyli uniwersaliów. Ale są to inne uniwersalia. Nie przynależą już do samego języka jako bytu autonomicznego, lecz są właściwością ogólnych zdolności ludzkiego umysłu, wśród których umiejętności postrzegania świata i przetwarzania informacji są jednymi z ważniejszych.

Zanim przejdziemy do przeglądu niektórych spośród dobrze znanych empirycznych ustaleń wskazujących zapotrzebowanie na niearystotelesowską teorię kategoryzacji, przypomnimy poglądy Wittgensteina i jego wpływ na badania kategoryzacji. Traktujemy Wittgensteinowską koncepcję podobieństw-

---

<sup>63</sup> Artowicz 1997, 1990.

wa rodzinnego wyrazów jako swego rodzaju przejście od językoznawstwa klasycznego do kognitywnego.

## WITTGENSTEINA KONCEPCJA PODOBIENSTWA RODZINNEGO

W *Dociekaniach filozoficznych* Wittgenstein podniósł problem, jak zdefiniować słowo *gra*. Po pierwsze zauważył, że rozmaite elementy kategorii GRA nie mają wspólnych własności, na podstawie których można odróżnić gry od czynności, które nie są grami. Granice kategorii są nieostre, jednakże fakt ten nie umniejsza komunikacyjnej użyteczności kategorii. Wbrew oczekiwaniom klasycznej teorii, kategoria GRA nie jest też ustrukturalizowana poprzez cechy kryterialne (dystynktywne), a raczej poprzez sieć krzyżujących się podobieństw. Oczywiście istnieją cechy typowo kojarzone z daną kategorią, przysługujące niektórym jej elementom i nie przysługujące innym. Jednakże nie ma atrybutów wspólnych wszystkim elementom i tylko im samym. Może też zdarzyć się, że niektóre elementy kategorii praktycznie nie będą miały nic wspólnego z pozostałymi. Oczywiście taki stan rzeczy nie jest specyficzny jedynie dla kategorii GRA. Znaczenia istotnej liczby wyrazów w różnych językach naturalnych są ustrukturalizowane w podobny sposób.

Aby opisać strukturę GRY, Wittgenstein posłużył się metaforą podobieństwa rodzinnego. „Przypatrz się np. kiedyś temu, co nazywamy <<grami>>. Chodzi mi tu o gry typu szachów, gry w karty, w piłkę, gry sportowe itd. Co jest im wszystkim wspólne? – Nie mów: <<Muszą mieć coś wspólnego, bo inaczej nie nazywałyby się grami>> – tylko patrz, czy mają coś wspólnego. – Gdy im się bowiem przypatrzysz, to nie dojrzysz wprawdzie niczego, co byłoby wszystkim wspólne, dostrzeżesz natomiast podobieństwa, pokrewieństwa – i to cały ich szereg. A więc jak się rzekło: nie myśl, lecz patrz! – Spójrz np. na gry typu szachów z ich rozmaitymi pokrewieństwami. Przejdź następnie do gier w karty: znajdziesz tu wiele odpowiedników tamtej klasy, ale też wiele rysów wspólnych znika, a pojawiają się inne. Gdy przechodzimy teraz do gier w piłkę, to niektóre cechy wspólne się zachowują, a wiele z nich się zatracą. – Czy każda z tych gier jest <<rozrywką>>? Porównajmy szachy z młynkiem. A może wszędzie istnieje wygrana i przegrana albo współzawodnictwo graczy? Przypomnijmy sobie pasjansa. W grach w piłkę istnieje wygrana i przegrana; gdy jednak dziecko rzuca piłkę o ścianę i znowu ją chwyta, to rysu tego już nie ma. Zobacz, jaką rolę grają sprawność i szczęście. I jak różna jest sprawność w szachach i w tenisie. Pomyśl dalej o grach typu korowodów tanecznych: jest tu element rozrywki, ale ileż to rysów charakterystycznych zniknęło! W ten oto sposób moglibyśmy przejść jeszcze przez wiele innych grup gier, obserwując pojawianie się i znikanie podobieństw. A wynik tych rozważań brzmi: Widzimy skomplikowaną siatkę zachodzących na siebie i krzyżujących się podobieństw; podobieństw w skali dużej i małej.



Podobieństw tych nie potrafię scharakteryzować lepiej niż jako «podobieństwa rodzinne», gdyż tak właśnie splatają się i krzyżują rozmaite podobieństwa członków jednej rodziny: wzrost, rysy twarzy, kolor oczu, chód, temperament itd. – Będę też mówić: «gry» tworzą rodzinę”<sup>64</sup>.

Dalej Wittgenstein omawia sposób uczenia się kategorii GRA. Ponieważ GRA nie jest ustrukturalizowana zgodnie z klasycznymi zasadami, kategorii tej nie można uczyć się jako koniunkcji tych cech dystynktywnych, które jednoznacznie odróżniają gry od nie-gier. Kategorii tej można uczyć się raczej na podstawie przykładów. Zamiast wyjaśniać komuś, czym jest gra, powinniśmy opisać mu gry, dodając, że *te i podobne rzeczy są nazywane grami*.

Według Taylora<sup>65</sup> Wittgenstein nic nie wspomina o wyborze przykładów – choć przypuszczalnie nikt nie uczyłby kategorii odwołując się tylko do przypadków granicznych; zaś szczegółowe wyartykułowanie pojęcia centralnego czy prototypowego okazu kategorii stało się zadaniem innych badaczy. Wittgenstein nie korzystał bowiem z pojęcia prototypu, utworzył i posługiwał się terminem pokrewnym – *podobieństwa rodzinnego* między pojęciami. Zdaniem Lakoffa<sup>66</sup> koncepcja Wittgensteina jest nie tylko jedną z najbardziej znaczących w rozwoju pojęcia podobieństwa rodzinnego, ale i pojęć centralności (okazu centralnego) i stopniowości. Podobny pogląd sformułowała Wierzbicka, pisząc „pisma Wittgensteina zawierają jedne z głębszych i najbardziej płodnych spostrzeżeń dotyczących kwestii semantycznych, jakie można znaleźć w literaturze przedmiotu”<sup>67</sup>. Wierzbicka podjęła jednak próbę sformułowania semantycznego inwariantu *gry*<sup>68</sup>, proponując następującą definicję:

- „*gry*
- (a) wiele rodzajów rzeczy, które ludzie robią
  - (b) ludzie mogą robić te rzeczy przez długi czas
  - (c) ludzie robią te rzeczy «dla przyjemności» (czyli dlatego, że chcą czuć coś dobrego)
  - (d) kiedy oni robią te rzeczy, chcą, żeby pewne rzeczy się zdarzyły
  - (e) gdyby ludzie nie robili tych rzeczy, nie chcieliby, żeby te rzeczy się zdarzyły
  - (f) kiedy oni robią te rzeczy, muszą wiedzieć, co mogą robić
  - (g) kiedy oni robią te rzeczy, muszą wiedzieć, czego nie mogą robić
  - (h) zanim ludzie będą robić te rzeczy, ktoś musi o tym powiedzieć”<sup>69</sup>.

---

<sup>64</sup> Wittgenstein 1972, s. 50-51.

<sup>65</sup> Taylor 1991, s. 33.

<sup>66</sup> Lakoff 1987, s. 14.

<sup>67</sup> Wierzbicka 1999, s. 37.

<sup>68</sup> W komentarzu do definicji gry (ang. *game*) Wierzbicka zwraca uwagę, że w języku niemieckim wyraz *Spiel* ma szerszy zakres użycia, odpowiadający mniej więcej angielskiemu *playing* (gra, zabawa).

<sup>69</sup> Wierzbicka 1999, s. 38.

Utworzona przez Wittgensteina metafora podobieństwa rodzinnego wydaje się być dobrym przejściem do pojęcia typowości, a ściślej kategorii prototypowej.

## KATEGORIE PROTOTYPOWE

Intuicja Wittgensteina o niewystarczalności klasycznej koncepcji kategoryzacji w odniesieniu do przynajmniej pewnych wyrazów z języka potocznego zyskała empiryczne potwierdzenie w licznych eksperymentach. Należą do nich między innymi prace prowadzone przez Williama Labova<sup>70</sup>. Labov badał kategoryzację językową naczyń gospodarstwa domowego, takich jak filiżanki, kubki, miseczki i wazony. Zastosował następującą procedurę. Przygotowano rysunki kształtów różnych naczyń. Następnie zostały one pokazane uczestnikom badania wraz z prośbą o nazwanie. Naczynia o kolistym przekroju w poziomie, zwężające się ku dołowi, których maksymalna szerokość była równa głębokości i które były zaopatrzone w uchwyt zostały jednogłośnie nazwane filiżanką. W miarę jak zwiększał się stosunek szerokości do głębokości, coraz częściej pojawiało się określenie miseczka. Wbrew oczekiwaniom nie było wyraźnej granicy między FILIŻANKĄ i MISECZKĄ, raczej jedna kategoria stopniowo wtapiała się w drugą. Pozbawienie naczyń uchwytu zmniejszało skłonność do kwalifikowania pokazywanych obiektów jako filiżanek, jednakże i tu wynik nie był jednoznaczny. Kategoryzację symulowano również, prosząc uczestników eksperymentu o wyobrażenie sobie naczyń wypełnionych różnymi rzeczami. Gdy myślano o kawie – wzrastała liczba odpowiedzi filiżanka, natomiast liczba ocen miseczka rosła na myśl o czymś do jedzenia a nie picia. Podobne efekty uzyskano, gdy głębokość zwiększała się bardziej niż szerokość. W tym wypadku kategoryzacja danego kształtu jako filiżanki ustąpiła miejsca kategoryzacji jako wazonu. Jeśli naczynia miały kształt bardziej cylindryczny niż zwężający się, częściej były kategoryzowane jako kubki. Na podstawie eksperymentu Labova sformułowano pewne wnioski, z których obecnie najważniejszy jest dla nas ten, że obiekty są kategoryzowane na podstawie ich atrybutów, które z kolei nie wydają się być binarnymi cechami dystynktywnymi w klasycznym rozumieniu. Weźmy na przykład pod uwagę stosunek szerokości kategoryzowanego obiektu do jego długości. Może on przybierać wartości z pewnego ciągłego przedziału. Wyniki eksperymentu Labova pokazują, że dla każdej z wyróżnionych kategorii charakterystyczna jest pewna wartość (lub zakres wartości) stosunku szerokości do długości. Przy kategoryzacji konkretnych obiektów jako FILIŻANEK, MISECZEK, WAZONÓW itd. istota tego procesu polegała nie na stwierdzeniu, czy dany obiekt posiada dany atrybut, czy też nie, lecz na ustaleniu jak dalece wymiary obiektu zbliżone są do jego wymiarów optymalnych. Nawet

<sup>70</sup> Kempton 1993. Cyt. za: Amerykańska 1993, s. 99-137.

obecność lub brak uchwytu nie są rozstrzygające. Po pierwsze, łatwo można sobie wyobrazić wiele możliwości pośrednich między pełnym uchwytem i niewielką wypukłością z boku naczynia. Po drugie, atrybuty – w przeciwieństwie do abstrakcyjnych cech dystynktywnych wyróżnianych w semantyce klasycznej – są własnościami obiektów rzeczywistego świata dostępnymi dla kompetentnych użytkowników języka na mocy znajomości otaczającej ich rzeczywistości. Atrybuty filiżanek, miseczek i wazonów obejmują charakterystyczny kształt, wielkość oraz materiał, z którego zrobione jest naczynie. Jednakże nie ograniczają się tylko do własności fizycznych. Labov wykazał, że jedno i to samo naczynie może być różnie kategoryzowane zależnie od tego, czy było używane do picia kawy, czy do jedzenia czegoś. Inaczej mówiąc, atrybuty mogą być niekiedy funkcjonalne (dotyczą użycia, do którego przeznaczony jest dany obiekt) lub mogą mieć interakcyjny charakter (dotyczą sposobu, w jaki ludzie posługują się obiektem). Wreszcie, atrybuty są związane nie z immanentnymi własnościami obiektu samego w sobie, lecz z rolą danego obiektu w określonej kulturze, na co wielokrotnie zwracała uwagę Wierzbicka.

Pojedynczy atrybut nie ma zasadniczego znaczenia w odróżnianiu jednej kategorii od drugiej. Obecność uchwytu lub funkcja naczynia jedynie zwiększa prawdopodobieństwo, że obiekt będzie kategoryzowany jako filiżanka. Jakkolwiek filiżanki zwykle mają uchwyty i są często używane do picia kawy, to obecność uchwytu i używanie do picia kawy nie są cechami definiującymi FILIŻANKĘ. Wszystkie filiżanki są też (potencjalnymi) pojemnikami. Bycie pojemnikiem nie może być jednak cechą definiującą FILIŻANKĘ. Wiele przedmiotów nie będących filiżankami również jest pojemnikami. Czym jest zatem to, co czyni filiżankę filiżanką a nie miseczką, czy wazonem? Co, w kategoriach arystotelesowskich, jest „istotą filiżanki”? Problem ten jest trudny do rozstrzygnięcia, przynajmniej na gruncie teorii Arystotelesa, choć równocześnie nie mamy problemów z uzmysławianiem czy rozpoznawaniem typowej filiżanki. Jakkolwiek FILIŻANKA może przenikać się z takimi kategoriami jak MISECZKA i WAZON, to istnieją pewne naczynia, które jedno myślnie i bezdyskusyjnie są kwalifikowane jako filiżanki.

Oczywista jest tu analogia z kategoriami kolorów. Tak jak pewne strefy w spektrum barw uchodzą za dobre, a nawet optymalne, przykłady CZERWONEGO, ŻÓLTEGO itd., tak też i naczynia gospodarstwa domowego okazują się podlegać kategoryzacji wokół dobrych, wyrazistych przykładów – prototypów – FILIŻANKI, MISECZKI, WAZONU itp. Te prototypy służą jako punkty odniesienia do kategoryzacji mniej wyraźnych przypadków.

Prototypy charakteryzują się bogactwem szczegółów często zdeterminowanych kulturowo, które – przy Arystotelesowskim podejściu – byłyby traktowane jako akcydentalne. I tak w naszej kulturze prototypowa filiżanka ma uchwyt, jest wykonana z porcelany, występuje ze spodkiem, ma określony kształt i typowy rozmiar; filiżanek używa się do picia herbaty lub kawy, często są kupowane w zestawach po sześć sztuk z dzbankiem. Żaden z tych atry-

butów nie decyduje jednak o przynależności do kategorii. Plastikowy pojemnik bez uchwyty i podstawki, taki, jaki dostaje się w automatach sprzedających kawę, jest ciągle filiżanką, choć nietypową.

Najbardziej rozległe i systematyczne empiryczne badania prototypów przeprowadziła Eleonor Rosch. Rosch zaczęła swoje prace od kategoryzacji kolorów, dochodząc między innymi do wniosku, że nazwy kolorów uzyskują swój zakres denotacyjny nie poprzez arbitralne ustalenie granic kategorii, lecz przez uogólnienie wychodzące od centralnych, prototypowych okazów. Późniejsze badania wykazały, że wiele innych naturalnych kategorii<sup>71</sup>, takich jak MEBLE czy PTAK, jest ustrukturalizowanych w podobny sposób. U Rosch, podobnie jak i u innych badaczy języka, zarówno generatywistów, jak strukturalistów i kognitywistów, podstawowym elementem procesu poznawczego jest dokonywanie porównań: szukanie podobieństw i różnic między rzeczami. Jest to klasyczna, Arystotelesowska zasada, będąca podstawą wszelkiej kategoryzacji. Rosch podkreśla natomiast, że procesy poznawcze są ograniczane przez cały szereg uniwersalnych i nieuniwersalnych czynników, takich jak indywidualne cechy osobowości, jednostkowe zdolności, poziom wykształcenia, system przekonań, motywacji, wartości, status społeczny, kontekst kulturowy itd., warunkujących percepcję. Wszelka percepcja jest z konieczności procesem wybiórczym: z jednej strony ograniczają ją fizyczne możliwości ludzkich zmysłów, z drugiej – nastawienie psychiczne. Ta wybiórczość percepcji znajduje bezpośrednie odbicie na płaszczyźnie struktury semantycznej znaków językowych, stając się w ten sposób istotnym elementem opisu zarówno gramatycznego, jak i języka jako całości.

Rosch badała strukturę naturalnych kategorii, pytając ankietowanych o ocenę pewnych rodzajów przedmiotów jako dobrych przykładów kategorii. Podana za Taylorem<sup>72</sup> tabela 1 odzwierciedla wyniki uzyskane od około 200 studentów amerykańskich college'ów, których poproszono o ocenę, do jakiego stopnia każdy z sześćdziesięciu przedmiotów domowego użytku można uznać za dobry przykład kategorii MEBLE (ang. FURNITURE). Ankietowani odpowiadali posługując się siedmiopunktową skalą od 1 (bardzo dobry przykład) poprzez 4 (umiarkowanie dobry przykład), do 7 (bardzo zły przykład lub nie jest przykładem w ogóle). W tabeli wymieniono obiekty według malejącego stopnia przynależności do kategorii. Rosch stwierdziła, że statystycznie biorąc, kolejność, w jakiej wymieniane są przedmioty w tabeli, jest wysoce wiarygodna. Wśród 200 ankietowanych odnotowano bardzo dużą zgodność, szczególnie w odniesieniu do przedmiotów wskazujących wyższy stopień przynależności do kategorii. Nawet różne podgrupy ankietowanych, na

---

<sup>71</sup> Przez naturalne kategorie Rosch rozumie pojęcia możliwe do oznaczenia słowami w językach naturalnych. Naturalne kategorie są przeciwstawiane kategoriom sztucznym, np. układom kropek lub ciągom liter i cyfr, często stosowanych przez psychologów w badaniach procesów uczenia się i kształtowania pojęć.

<sup>72</sup> Taylor 1991, s. 44.

Tabela<sup>73</sup> 1. Reprezentatywność sześćdziesięciu okazów kategorii MEBLE

| Okaz                   | Ranga | Ocena trafności | Okaz               | Ranga | Ocena trafności |
|------------------------|-------|-----------------|--------------------|-------|-----------------|
| krzesło                | 1,5   | 1,04            | lampa              | 31    | 2,94            |
| sofa                   | 1,5   | 1,04            | taboret            | 32    | 3,13            |
| kanapa                 | 3,5   | 1,10            | klęcznik           | 33    | 3,43            |
| stół                   | 3,5   | 1,10            | szuflada           | 34    | 3,63            |
| fotel                  | 5     | 1,33            | pianino            | 35    | 3,64            |
| szafa                  | 6,5   | 1,37            | poduszka           | 36    | 3,70            |
| fotel-bujak            | 6,5   | 1,37            | stojak na gazety   | 37    | 4,14            |
| stolik kawowy          | 8     | 1,38            | hi-fi              | 38    | 4,25            |
| kołyska                | 9     | 1,42            | szafka na naczynia | 39    | 4,27            |
| fotel dwu-<br>osobowy  | 10    | 1,44            | stereo             | 40    | 4,32            |
| komoda                 | 11    | 1,48            | lustro             | 41    | 4,39            |
| biurko                 | 12    | 1,54            | telewizor          | 42    | 4,41            |
| łóżko                  | 13    | 1,58            | barek              | 43    | 4,46            |
| sekretarzyk            | 14    | 1,59            | półka              | 44    | 4,52            |
| tapczan                | 15,5  | 1,61            | chodnik            | 45    | 5,00            |
| pomocnik<br>(stolik)   | 15,5  | 1,61            | poduszka do spania | 46    | 5,03            |
| materac na<br>stelażu  | 17    | 1,70            | kosz na śmieci     | 47    | 5,34            |
| nocna szafka           | 18    | 1,83            | radio              | 48    | 5,37            |
| kufer                  | 19    | 1,98            | maszyna do szycia  | 49    | 5,39            |
| skrzynia (kufer)       | 20    | 2,11            | kuchnia            | 50    | 5,40            |
| cedrowy<br>serwantka   | 21    | 2,13            | łada               | 51    | 5,44            |
| regał<br>książkowy     | 22    | 2,15            | zegar              | 52    | 5,48            |
| kozetka                | 23    | 2,17            | kotary             | 53    | 5,67            |
| szezłąg                | 24    | 2,26            | lodówka            | 54    | 5,70            |
| otomana                | 25    | 2,43            | obraz              | 55    | 5,75            |
| podnózek               | 26    | 2,45            | szafa wnękowa      | 56    | 5,95            |
| gablotka               | 27    | 2,49            | wazon              | 57    | 6,23            |
| szafka na<br>porcelanę | 28    | 2,59            | popielniczka       | 58    | 6,35            |
| ława                   | 29    | 2,77            | wentylator         | 59    | 6,49            |
| bufet                  | 30    | 2,89            | telefon            | 60    | 6,68            |

<sup>73</sup> Mając na uwadze naturalne mankamenty przekładu, zdecydowaliśmy się zamieścić również oryginalne wersje tabel 1 i 2.

Tabla 1a. Goodness-of-example ratings for sixty members of the category FURNITURE<sup>74</sup>

| Member           | Rank | Specific score | Member         | Rank | Specific score |
|------------------|------|----------------|----------------|------|----------------|
| chair            | 1,5  | 1,04           | lamp           | 31   | 2,94           |
| sofa             | 1,5  | 1,04           | stool          | 32   | 3,13           |
| couch            | 3,5  | 1,10           | hassock        | 33   | 3,43           |
| table            | 3,5  | 1,10           | drawers        | 34   | 3,63           |
| easy chair       | 5    | 1,33           | piano          | 35   | 3,64           |
| dresser          | 6,5  | 1,37           | cushion        | 36   | 3,70           |
| rocking-chair    | 6,5  | 1,37           | magazine rack  | 37   | 4,14           |
| coffe table      | 8    | 1,38           | hi-fi          | 38   | 4,25           |
| rocker           | 9    | 1,42           | cupboard       | 39   | 4,27           |
| love seat        | 10   | 1,44           | stereo         | 40   | 4,32           |
| chest of drawers | 11   | 1,48           | mirror         | 41   | 4,39           |
| desk             | 12   | 1,54           | television     | 42   | 4,41           |
| bed              | 13   | 1,58           | bar            | 43   | 4,46           |
| bureau           | 14   | 1,59           | shelf          | 44   | 4,52           |
| davenport        | 15,5 | 1,61           | rug            | 45   | 5,00           |
| end table        | 15,5 | 1,61           | pillow         | 46   | 5,03           |
| divan            | 17   | 1,70           | wastebasket    | 47   | 5,34           |
| night table      | 18   | 1,83           | radio          | 48   | 5,37           |
| chest            | 19   | 1,98           | sewing machine | 49   | 5,39           |
| cedar chest      | 20   | 2,11           | stove          | 50   | 5,40           |
| vanity           | 21   | 2,13           | counter        | 51   | 5,44           |
| bookcase         | 22   | 2,15           | clock          | 52   | 5,48           |
| longue           | 23   | 2,17           | drapes         | 53   | 5,67           |
| chaise longue    | 24   | 2,26           | refrigerator   | 54   | 5,70           |
| ottoman          | 25   | 2,43           | picture        | 55   | 5,75           |
| footstool        | 26   | 2,45           | closet         | 56   | 5,95           |
| cabinet          | 27   | 2,49           | vase           | 57   | 6,23           |
| china closet     | 28   | 2,59           | ashtray        | 58   | 6,35           |
| bench            | 29   | 2,77           | fan            | 59   | 6,49           |
| buffet           | 30   | 2,89           | telephone      | 60   | 6,68           |

<sup>74</sup> Taylor 1991, s. 44.

przykład tych, którzy mieszkali głównie na wschodnim wybrzeżu Stanów Zjednoczonych, i tych, którzy pochodzili z wybrzeża zachodniego, dostarczyły porównywalnych odpowiedzi. Z klasycznego punktu widzenia pytanie o to, w jakim zakresie dana rzecz należy do kategorii jest pozbawione sensu, gdyż rzecz może jedynie należeć lub nie należeć do kategorii. Rosch wykazała jednak, że pytanie o stopień przynależności do kategorii jest problemem realnym.

Innym ważnym efektem badań Rosch było stwierdzenie, że stopień przynależności do kategorii wpływa na tempo weryfikacji sądów typu  *pewien x jest pewnym Y*. Zweryfikowanie, że drozd (centralny okaz kategorii PTAK w badaniach Rosch) jest ptakiem, zajmuje mniej czasu niż zweryfikowanie tego, że kaczka jest ptakiem. Efekt ten jest bardziej wyrazisty przy badaniu dzieci, co prawdopodobnie wynika z niepełnego przyswojenia przez dzieci mniej typowych czy marginalnych okazów kategorii. Przeprowadzono następujący eksperyment. Na ekranie ukazywały się dwa wyrazy. Ankietowany musiał określić, możliwie najszybciej, czy te dwa wyrazy są takie same, czy różne. Prezentacja każdej pary była poprzedzona prezentacją nazwy kategorii, na przykład para *krzesło-krzesło* była poprzedzona MEBLE. Jeśli prezentowane wyrazy były dobrymi okazami kategorii, to czas odpowiedzi był krótszy. Jeśli zaś były słabymi okazami kategorii, na przykład *piec* jako reprezentant kategorii MEBLE, wówczas czas odpowiedzi był dłuższy. Wynik ten pozwala przypuszczać, że nazwa kategorii aktywizuje nazwy bardziej typowych okazów kategorii i w jakimś stopniu dezaktywizuje okazy marginalne. Uderzająca okazała się również korelacja między stopniem przynależności do kategorii i częstością oraz kolejnością, w jakiej wymieniane są okazy kategorii.

Pewnego uzasadnienia, a zarazem wyjaśnienia procesu kategoryzacji przez prototyp, dostarcza analiza procesów percepcyjnych. Przyjmuje się, że prototyp jest konstrukcją poznawczą, dającą się odnieść do jakiegoś fragmentu otaczającej rzeczywistości lub w jakiś sposób z takim fragmentem skorelować. Jest on więc niejako symbolem danej kategorii, poznawczym „punktem odniesienia”. Funkcję prototypu może pełnić najlepszy (w naszym przekonaniu) egzemplarz będący przedstawicielem danej kategorii, ale może to być także schemat konceptualny, tj. pewien zestaw cech, które (w naszym przekonaniu) powinny charakteryzować dobre elementy danej kategorii.

Granice między kategoriami prototypowymi są nieostre, a poszczególne kategorie nakładają się na siebie. Ponieważ w obrębie każdej kategorii znajdują się zarówno elementy, które ludzie uznają za dobre przykłady, jak i takie, które są uznawane za elementy gorsze, im dalej te drugie znajdują się od przedstawicieli prototypowych (centralnych), tym większa możliwość nałożenia się na siebie kilku kategorii. Wbrew założeniom kategoryzacji klasycznej, aby dokonać przyporządkowania obiektu do kategorii, wcale nie musimy wiedzieć, gdzie dokładnie przebiegają granice między kategoriami. Nie musimy umieć zdefiniować na przykład *samochodu* i *windy*. Wystarczy, jeśli potrafimy od siebie odróżnić samochód i windę.

Wiązki atrybutów przypisywanych okazom uznawanym za najlepszych (prototypowych) reprezentantów swoich kategorii tworzą „całościowe struktury pojęciowe”<sup>75</sup>. Struktury takie powstają w codziennym doświadczeniu – w odróżnieniu od struktur pojęciowych tworzonych na użytek nauki. Nie opisują rzeczywistego świata, lecz są jego cząstkowymi wyobrażeniami. Nie muszą być ani prawdziwe, ani kompletne, natomiast ze swej istoty są subiektywne<sup>76</sup>. Kategorie prototypowe są kategoriami ogólnej wiedzy o świecie i języka potocznego, podczas gdy kategorie klasyczne są raczej kategoriami nauki. W językach naturalnych występują oba rodzaje kategorii, przy czym pewne pojęcia są definiowane tylko w klasycznych kategoriach naukowych, pewne tylko w kategoriach potocznych, jeszcze inne zaś zarówno w kategoriach naukowych, jak i potocznych.

Interesująco teorię prototypów skomentował Przemysław Łozowski. „Prototyp jest więc czymś, od czego cały proces kategoryzacji miałby się zaczynać – gotową matrycą, która określałaby stopień przynależności do danej kategorii jej mniej lub bardziej prototypowych elementów. W tej sytuacji sama kategoria, z przypadkami centralnymi i peryferyjnymi, jawiłaby się jako wysoce dynamiczna struktura, w której cały czas zachodzi proces ścierania się (modelowania) atrybutów w oparciu o stały wzorzec. Prototyp natomiast byłby idealnym odwzorowaniem tego, co kategoria ma przedstawiać – niezmiennym jądrem kategorii. Co więcej, ponieważ to właśnie elementy peryferyjne są zazwyczaj <<kwestionowane>> w procesie poznawczym (w wyniku czego następują próby usprawiedliwienia ich obecności w kategorii poprzez odniesienie do prototypu), jedynie one są też w stanie zburzyć obraz kategorii zamkniętych. Wydaje się jednak, że intencja Rosch jest nieco inna. Równie ważna i interesująca w jej pracach jest ta część łańcucha kategoryzacji, która obejmuje proces kształtowania się prototypu. Prototyp nie jest bowiem wartością statyczną i nie może być rozumiany jako zestaw warunków gwarantujących przynależność do kategorii. Gdyby tak było istotnie, na początku listy rankingowej najbardziej prototypowych przykładów kategorii MEBLE (...) nie znalazłyby się dwa różne przedmioty – *chair* i *sofa* – mające identyczną liczbę uzyskanych punktów. Który z nich miałby ewentualnie służyć jako punkt odniesienia? Czy każdy z nich jest prototypem kategorii, czy dopiero razem dają pełny obraz tego, czym jest ich kategoria?”<sup>77</sup>. W udzieleniu odpowiedzi na pytanie *Czym jest prototyp?* nie pomaga zresztą sama Rosch, która przestrzega przed takim rozumieniem prototypu, według którego „miałby on oznaczać albo jednostkowy element kategorii, albo pewną strukturę mentalną”<sup>78</sup>.

<sup>75</sup> Tabakowska 1995, s. 43.

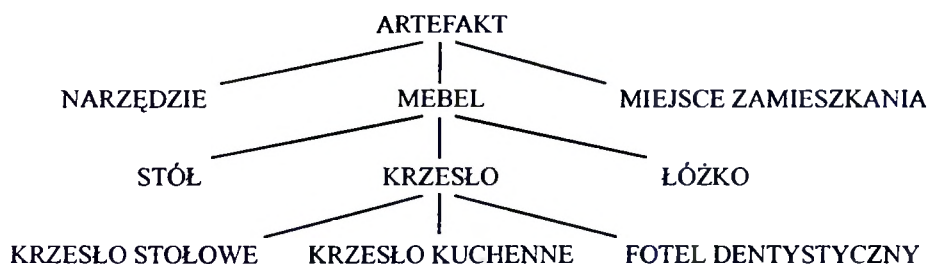
<sup>76</sup> „Wszelka wiedza zależy bowiem od tego, kim jest wiedzący”. Tabakowska 1995, s. 45.

<sup>77</sup> Łozowski 1994, s. 245.

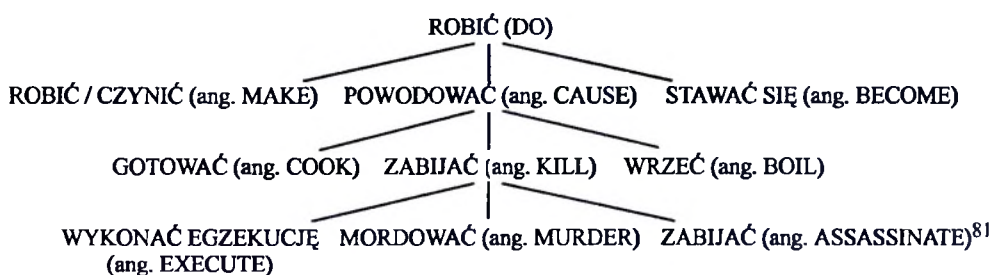
<sup>78</sup> Ibidem s. 246.



Prototypowość w ujęciu badanym przez Rosch jest ściśle związana z tym, co bywa nazywane dwiema osiami kategoryzacji (rys. 1 i 2)<sup>79</sup>. KRZESŁO, MEBLE, ARTEFAKT, RZECZ to kategorie czterech różnych poziomów. Kategoria KRZESŁO zawiera się w nadrzędnej kategorii MEBLE, która z kolei zawiera się w kategorii ARTEFAKT. Z drugiej strony KRZESŁO KUCHENNE jest podrzędnym okazem kategorii KRZESŁO. Te różne poziomy kategoryzacji tworzą oś pionową, zwaną osią inkluzji. Według Rosch kategorie uporządkowane są według zasady inkluzji, czyli zawierania. Na osi poziomej kategorie uporządkowane są według zasady segmentacji. „Istnieje zatem ścisła gradacja poziomów abstrakcji w myśl zasady, że im bardziej obszerna kategoria, tym wyższy poziom abstrakcji w płaszczyźnie pionowej. W płaszczyźnie poziomej natomiast występuje gradacja ciężaru gatunkowego przypisywanego poszczególnym reprezentantom danej kategorii”<sup>80</sup>. I tak, NARZĘDZIE, MEBLE, MIESZKANIE – wszystkie są przykładami kategorii ARTEFAKT; STÓŁ, KRZESŁO, ŁÓŻKO są okazami MEBLE, podczas gdy KRZESŁO STOŁOWE, KRZESŁO KUCHENNE, FOTEL DENTYSTYCZNY są okazami KRZESŁO. Na bardzo niskim poziomie kategorii znajdują się nie tyle kategorie, co raczej pojedyncze okazy. Zniszczone krzesło kuchenne, stojące w mojej kuchni może być okazem kategorii KRZESŁO KUCHENNE.



Rys. 1. Dwie osie kategoryzacji: kategorie nominalne



Rys. 2. Dwie osie kategoryzacji: kategorie werbalne

<sup>79</sup> Rysunki 1 i 2 są przejęte z: Taylor 1991, s. 47-48.

<sup>80</sup> Łozowski 1994, s. 247.

<sup>81</sup> Różnica między 'murder' i 'assassinate' wyraża się głównie w tym, że obiektem czynu (morderstwa, zabójstwa) nazywanego *assassinate* są najczęściej osoby znane, odgrywające ważne role w życiu publicznym oraz że zwykle robione jest ono dla piciny. Słowniki porównawcze angielsko-polskie najczęściej wskazują wyrazy *mordować*, *zabijać*, *zamach* jako ekwiwalenty *assassinate*.

Z pojęciem osi pionowej kategoryzacji związane jest, wywodzące się od Rosch, wyróżnianie istotnego kognitywnie i językowo poziomu kategoryzacji zwanego poziomem podstawowym. Podstawowe poziomy kategoryzacji to struktury uznane przez użytkowników języka za najbardziej użyteczne. Stanowią one punkt wyjścia do tworzenia dalszych kategorii – poziomów nadrzędnych i podrzędnych. Według Rosch w leksykonie każdego języka naturalnego status terminów poziomu podstawowego uzyskuje to pięć hierarchii, które okazuje się najużyteczniejsze w codziennej komunikacji, najskuteczniejsze i najbardziej ekonomiczne w użyciu. To właśnie na podstawowym poziomie kategoryzacji ludzie konceptualizują rzeczy jako wyobrażenia perceptualne i funkcjonalne. Spróbujmy, tłumaczy Rosch, przedstawić sobie wizualnie lub naszkicować na przykład mebel. Zadanie wydaje się niewykonalne bez uzyskania odpowiedzi na pytania: Jaki rodzaj mebla? Stół, krzesło czy łóżko? Inaczej sytuacja wyglądałaby w przypadku prośby o naszkicowanie obrazu krzesła, choć i wówczas równie zasadne byłoby pytanie: Jaki rodzaj krzesła? Podobnie rzecz się ma z próbą opisu postępowania z jakimś nieokreślonym meblem, czy też nazwania elementów, z których składa się jakiś nieokreślony mebel. Nie mamy natomiast kłopotów z opisem ruchów wykonywanych przy posługiwaniu się krzesłem lub z nazwaniem części, z których składa się krzesło. Nie jest zatem niczym zaskakującym, że to właśnie podstawowy poziom klasyfikacji jest tym poziomem, na którym ludzie zwykle mówią o rzeczywistości. „Gdyby cudzoziemiec wskazał przedmiot, na którym właśnie siedzę i zapytał <<Jak to się nazywa po angielsku>>, z pewnością odpowiedziałbym <<To jest krzesło>>. Nie odpowiedziałbym <<To jest artefakt>> lub <<To jest pojedynczy mebel>>, nawet gdyby te alternatywne odpowiedzi były jednakowo poprawne”<sup>82</sup>. Pojęcie podstawowego poziomu wyrażenia jest najczęściej odnoszone do kategorii nominalnych. Jednakże można je stosować również w odniesieniu do pojęć innego rodzaju. Rosch<sup>83</sup> przedstawia dowody sugerujące, że wyrażenia takie jak *zmywać* (ang. *do washing up*) i *myć zęby* (ang. *brush one's teeth*) mają status wyrażenia poziomu podstawowego, podczas gdy *wykonywać domowe zajęcia* (ang. *do the household chores*) i *wyciskać pastę do zębów na szczoteczkę* (ang. *squeese toothpaste onto the brush*) mają odpowiednio status nadrzędny i podrzędny.

„Na gruncie czysto formalnym, wewnątrzjęzykowym wyrażenia podstawowego poziomu często dają się odróżnić od wyrażenia niepodstawowych. Cechą dodatkową wyrażenia podstawowego poziomu, oprócz wysokiej frekwencji, jest ich krótkość i prostota strukturalna (są najczęściej monomorfemiczne). Wyrażenia poziomu niższego niż podstawowy są często złożeniami wyrażenia poziomu podstawowego oraz wyrażenia modyfikującego (np. *kitchen chair*). Wyrażenia powyżej poziomu podstawowego są niekiedy na swój sposób anomaliami (np. ang. *furniture* jest morfosyntaktycznie nietypowe

<sup>82</sup> Taylor 1991, s. 49.

<sup>83</sup> Cognition 1978, s. 43-46.

jako niepoliczalne, tj. po angielsku nie można powiedzieć a *furniture* lub *furnitures*). Często zdarza się też, że wyrażenia nadrzędne nie występują w słowniku danego języka. Na przykład angielski ma wiele nazw kolorów, jednakże nie ma nazwy nadrzędnej dla kolorów. *Coloured* nie jest tu odpowiedni, ponieważ odnosi się jedynie do przypadków barw i wyklucza *black*, *white* i *gray*. Często, gdy potrzeba wyrażen nadrzędnych jest silnie odczuwana, wymyśla się terminy naukowe (takie jak *sibling* jako nadrzędne dla *brother* i *sister*), czy też niezręczne omówienia”<sup>84</sup>.

Co decyduje o uprzywilejowanym statusie wyrażen podstawowego poziomu? Psychologowie i psycholingwiści sugerują, że odpowiedzi należy poszukiwać w użyteczności kategorii. Kategoryzacja umożliwia ograniczenie do uchwytnych rozmiarów nieskończenie wielu odmian rzeczy w rzeczywistości. Wiąże się między innymi ze sposobem, w jaki ludzie dokonują segmentacji rzeczywistości do takiego poziomu, na którym bodźce nietożsame (nie-identyczne) mogą uchodzić za ekwiwalentne. Kategoria spełnia tę funkcję na mocy takiej własności, że znając kategorię, do której należy obiekt, zna się tym samym tyle atrybutów rzeczy, ile możliwe. Dzięki tej samej własności możliwe jest wykluczenie wielu innych atrybutów. Możliwość funkcjonowania kategorii w ten sposób opiera się na fakcie, że atrybuty nie tylko nie występują w rzeczywistości przypadkowo, lecz wykazują również tendencję do wzajemnej korelacji, pozytywnej lub negatywnej. Inaczej mówiąc, występowanie w rzeczywistości atrybutu A oznacza prawdopodobieństwo powiązania z wystąpieniem atrybutu B i brakiem atrybutu C. Jeśli wiemy, że jakiś obiekt jest opierzony, ma skrzydła i potrafi fruwać, to możemy stwierdzić z pewnym przekonaniem, że również znosi jajka. Z drugiej strony, posiadanie futra i ogona są atrybutami nie kojarzonymi ze znoszeniem jajek i zdolnością fruwania. Arystotelesowski model kategorii zakładał doskonałą korelację między cechami dystynktywnymi w ramach kategorii. W ujęciu tym znajomość kategorii, do której należy rzecz, oznaczała absolutną pewność, że współwystąpią cechy będące warunkami niezbędnymi uznania przynależności do danej kategorii. Doświadczenie jednakże uczy, że w rzeczywistości nie zawsze ma się do czynienia z taką idealną korelacją. Istnieją filizanki bez uchwyty (chińskie filizanki), ptaki, które nie fruwią (pingwiny), koty bez ogonów (manksy – koty z wyspy Man) itd.

Zdaniem Rosch to właśnie kategorie podstawowego poziomu najpełniej realizują rzeczywiste korelacje atrybutów zachodzące w świecie zewnętrznym. Wyrażenia podstawowego poziomu rozczłonkują rzeczywistość na kategorie maksymalnie informatywne, zwiększając zarówno liczbę atrybutów przysługujących okazom kategorii, jak i zmniejszając liczbę atrybutów dzielonych z okazami innych kategorii. Rozważmy jeszcze raz kategorię MEBLE. Patrząc na tabelę 1, trudno byłoby wskazać atrybuty definiujące tej

---

<sup>84</sup> Taylor 1991, s. 49.

kategori. Nawet jeśli ograniczymy się do około trzydziestu typowych okazów kategorii, nie pojawiają się atrybuty jednoznacznie odróżniające artefakty „meblowe” od innych artefaktów domowego użytku. W przeciwieństwie do założeń teorii klasycznej o kategorii MEBLE prawdopodobnie myślimy nie w kategoriach koniunkcji cech dystynktywnych, lecz raczej poprzez wskazywanie jej bardziej typowych okazów, takich jak: stoły, łóżka, kanapy, krzesła itp. W przypadku kategorii podstawowej, jak np. KRZESŁO sytuacja jest inna. Krzesła mają dosyć wiele wspólnych atrybutów. Co więcej, to, co wspólne dla krzesła, nie musi przysługiwać łóżkom i stołom. Okazy kategorii na kolejnym, najniższym poziomie, np. KRZESŁO KUCHENNE, również dzielą dużą liczbę wspólnych atrybutów. Jednakże wiele spośród atrybutów krzesła kuchennych przysługuje innym rodzajom krzesła, np. krzesłom biurowym. O ile KRZESŁO KUCHENNE maksymalizuje liczbę atrybutów przysługującym okazom kategorii, o tyle nie odróżnia ich maksymalnie od kategorii na tym samym poziomie.

To właśnie w tym miejscu pojęcie poziomu podstawowego zajął się z prototypową strukturą kategorii. Kategorie mają zwykle rozmyte granice i mogą nakładać się wzajemnie; niektóre atrybuty mogą przysługiwać zaledwie kilku okazom kategorii; mogą wystąpić nawet kategorie bez atrybutów przysługujących wszystkim okazom. W celu utrzymania maksymalnego rozróżnienia kategorii, a więc w celu zapewnienia ich maksymalnej informatywności, musimy się skoncentrować na podstawowym poziomie kategoryzacji, a dokładniej na bardziej centralnych okazach kategorii podstawowego poziomu. Jeśli przyjąć za rzeczywistą stopniowaną przynależność do kategorii, to powstaje pytanie, dlaczego pewne przykłady kategorii uzyskują uprzywilejowany status okazów prototypowych, podczas gdy inne są przykładami marginalnymi. Jak to się dzieje, że krzesła, kanapy i stoły, a nie lustra, półki i zegary są prototypowymi meblami? Rosch rozważa kilka możliwych odpowiedzi na to pytanie. Dla pewnej ograniczonej liczby kategorii prototypowość jest z dużym prawdopodobieństwem skutkiem wrodzonych własności systemu percepcji człowieka. Na przykład, jest bardzo prawdopodobne, że prototypowy charakter centralnych kolorów ma oparcie w neurofizjologii percepcji barw. Inaczej mówiąc, istnienie kategorii kolorów w pewnym sensie poprzedza ich językowe kodowanie, choć nie można wykluczyć, że centralne znaczenie nazw pewnych kolorów może być skutkiem znaczenia pewnych obiektów w danej kulturze. Są dowody, że pewne formy geometryczne (tzw. dobre formy w psychologii przedstawię, np. koło, kwadrat, trójkąt) i pewne orientacje przestrzenne (np. pionowa i pozioma bardziej niż ukośna), podobnie jak centralne kolory, percepcyjnie są bardziej wyraziste, a przez to również uzyskują status prototypowy. Przypuszczalnie kategorie tego rodzaju będzie stosunkowo niewiele. Trudno jednak pytać o neurofizjologiczne podstawy per-

cepcji prototypowych filiżanek i mebli, ponieważ są to kategorie sztuczne, tzn. będące produktami środowiska kulturowego. Prototypowa struktura tego typu kategorii wymaga szukania innych wyjaśnień.

Jedno z możliwych wyjaśnień polegałoby na tym, że pewne okazy kategorii uzyskują prototypowy status, ponieważ mamy z nimi częściej do czynienia. W swoich empirycznych badaniach Rosch dokładnie badała frekwencję słów i doszła do przekonania, że stopień przynależności do kategorii jest niezależny od częstości wystąpień nazw okazów. Więcej, Rosch przestrzega, aby być ostrożnym we frekwencyjnym wyjaśnieniu prototypowości. Jej zdaniem wrażenie o wyższej frekwencji wystąpień okazów prototypowych może być symptomem prototypowości, a nie jej przyczyną. Respondenci proszeni o nazwanie przykładów kategorii byli skłonni do wymieniania w pierwszej kolejności okazów prototypowych. Jeśli zatem rozważając kategorię MEBLE, myślimy o stołach i krzesłach, a nie o lustrach i zegarach, to możemy wyprowadzić – być może błędny – wniosek, że częściej posługujemy się stołami i krzesłami niż lustrami i zegarami ściennymi. Warto wspomnieć, że to właśnie w tym duchu Rosch tłumaczy to, co nazywa „efektem dobrych dawnych czasów” (ang. *good old days effect*). Wspominając wcześniejsze zdarzenia z naszego życia, mamy skłonność do przesady – zawyżania częstości prototypowych i obniżania częstości nieprototypowych składników sytuacji. Dlatego przeszłość jest zapamiętywana jako przyjemna (lub nieprzyjemna, zależnie od przypadku)<sup>85</sup>. Innym czynnikiem proponowanym przez Rosch jest kolejność uczenia się. Na przykład nazwy przedmiotów, jakimi posługują się rodzice wobec małych dzieci, pochodzą głównie z poziomu podstawowego, i te nazwy dzieci przyswajają w pierwszej kolejności<sup>86</sup>. Jeszcze inne wyjaśnienie mogłoby polegać na tym, że w pewnych przypadkach prototyp może obejmować średnie wartości atrybutów zmiennej. Odwrotnie, pewne atrybuty mogą być szczególnie wyraziste, np. ponieważ są szczególnie ważne w społeczeństwie, czego skutkiem jest skupianie się tych atrybutów w prototypach.

Inaczej na problem hierarchii kategorii i pojęcie poziomu podstawowego patrzy Wierzbicka. Jej zdaniem „kategoryzacja taksonomiczna odgrywa rzeczywiście pewną rolę w myśleniu ludzkim odbitym w języku naturalnym, ale nie dominuje w tym myśleniu tak, jak to się przyjęło sądzić. Ważną rolę grają też inne zasady klasyfikacji, w niektórych partiach leksyki – nawet rolę ważniejszą. (...) Wiele domniemyanych taksonomii – zarówno w dziedzinie pojęć biologicznych, jak też w dziedzinie artefaktów – to w gruncie rzeczy nie są żadne taksonomie i że problem wyrazistości psychologicznej nie może

---

<sup>85</sup> Ten rodzaj wyjaśnień można przypisać faktowi, że dziecięce rysunki zwykle przedstawiają trawę jako nieczłowieczoną zieloną, a niebo jako ciągły błękit, choć w różnych warunkach niebo częściej jest szare niż niebieskie, zaś trawa przez większą część roku może być brunatna. Rysunki dziecięce przedstawiają kolor prototypowej trawy i prototypowego nieba. Trawa jest prototypowo zielona, co nie oznacza tego samego, co stwierdzenie, że trawa jest zwykle zielona.

<sup>86</sup> Kurcz 1995, s. 228.

zostać rozwiązany, dopóki rzeczywiste kategorie taksonomiczne nie zostaną oddzielone od kategorii nietaksonomicznych<sup>87</sup>. Prawdopodobnie różnica sądów jest związana z tym, że Wierzbicka rozważa problem kategorii i inkluzji z punktu widzenia semantyki językoznawczej, podczas gdy wcześniej prezentowane koncepcje były pracami antropologów, psychologów i psycholingwistów.

Wierzbicka stara się wprowadzić ostre rozróżnienie między relacją pojęciową określoną przez nią jako „rodzaj czegoś” (w płaszczyźnie obiektowej odpowiada jej superkategoria będąca rodzajem rzeczy) i relacją referencjalną inkluzji. „Kiedy definiujemy *psa* jako „rodzaj zwierzęcia”, *papugę* jako „rodzaj ptaka”, *różę* jako „rodzaj kwiatu”, *wiąz* jako „rodzaj drzewa”, zakładamy, że pojęciowo pies jest przede wszystkim „rodzajem zwierzęcia”, papuga „rodzajem ptaka” itd. i że wszystkie inne cechy definicyjne są podporządkowane tej właśnie cesze. (...) Pojęcie, które służy do identyfikowania rodzaju rzeczy, zastępuje coś, co ma pewną charakterystykę percepcyjną i może być wyobrażone albo narysowane. Tak więc ktoś może narysować nie tylko papugę, ale i jakiegoś nieokreślonego ptaka, nie tylko różę, ale także jakiś kwiat, nie tylko drzewo dębu, ale także drzewo w ogóle<sup>88</sup>. Wierzbicka, doceniając wkład Rosch w stworzenie i objaśnianie pojęcia poziomów podstawowych, tak między innymi to komentuje. „(...) wszystkie pojęcia taksonomiczne (które odnoszą się do rodzaju rzeczy) mogą być przedstawiane obrazowo niezależnie od tego, czy są, czy nie są podstawowe w sensie sprecyzowanym przez E. Rosch. Jeśli pojęcia najbardziej wyraziste psychologicznie mogą być przedstawione obrazowo, to jest tak nie dlatego, że są one psychologicznie najbardziej wyraziste, ale dlatego, że są to pojęcia odnoszące się do rodzaju rzeczy. Z drugiej strony jest możliwe, że tylko takie pojęcia, które odnoszą się do rodzaju rzeczy, mogą być psychologicznie wyraziste. Ale aby to zbadać, musimy wpięrcw ustalić, które pojęcia są taksonomiczne, a które nie<sup>89</sup>.”

Wierzbicka wyróżniła trzy rodzaje pojęć: taksonomiczne, tj. odnoszące się do specyficznych rodzajów rzeczy; funkcjonalne, odnoszące się do rzeczy pewnego rodzaju, mających specyficzny rodzaj funkcji; kolektywne, tj. oparte na sąsiedztwie, przyległości (ang. *contiguity*). Pojęcia taksonomiczne identyfikują rodzaj rzeczy, zastępują coś, co ma pewną charakterystykę percepcyjną i może być wyobrażone czy narysowane. Można zatem narysować nie tylko papugę, ale i jakiegoś nieokreślonego ptaka, nie tylko różę, wiaz itd., ale i jakiś kwiat, drzewo itd. Nie można jednak narysować zabawki, broni, pojemnika czy sprzętu w ogóle, ponieważ pojęcia te są funkcjonalne, nie implikują właściwości zmysłowych i dlatego nie mogą być przedstawione obrazowo. „Pojęcia funkcjonalne nie są ograniczone do dziedziny artefaktów. Także zwierzęta

<sup>87</sup> Wierzbicka 1999, s. 83-84.

<sup>88</sup> Ibidem s. 85.

<sup>89</sup> Ibidem s. 86.

i rośliny mogą być konceptualizowane z perspektywy funkcjonalnej i takie ich widzenie może zostać utrwalone w słownictwie. Dobrymi przykładami są słowa takie, jak *ulubieniec* [*pet*] i *chwast* [*weed*]<sup>90</sup>. Pojęcia odnoszące się do rodzaju rzeczy oparte są zasadniczo na podobieństwie między denotatami. Pojęcia funkcjonalne oparte są na tożsamości funkcji i nie implikują podobieństwa między denotatami. Trzeci typ pojęć nie opiera się ani na podobieństwie, ani na funkcji, ale na sąsiedztwie. Należą tu przede wszystkim pojęcia części pewnych całości. „Antropologowie (...) często odróżniają – i słusznie – to, co nazywają klasami (taxonomies) (systemy rodzajów), od tego, co nazywają systemami opartymi na pojęciu części składowych (partonomies) (systemy części). Tak więc *nos* nie jest rodzajem rzeczy (która wystaje w środku twarzy), ale częścią twarzy (która wystaje w środku twarzy itd.); *głowa* nie jest rodzajem rzeczy (która jest ponad innymi częściami ciała), ale – częścią ciała (która jest ponad innymi częściami ciała itd.). (...) Pojęcia oparte raczej na ciągłości niż na podobieństwie istnieją także na poziomie superkategorii. Podczas gdy na niższych poziomach, takich jak poziom słów nazywających części ciała, odróżniono już partonomie od taksonomii, na poziomach wyższych systemy oparte raczej na ciągłości niż na podobieństwie są generalnie mylone z taksonomiami<sup>91</sup>. Do pojęć kolektywnych opartych na sąsiedztwie Wierzbicka zaliczyła ponadto singularia tantum (np. *odzież*), pluraria tantum (np. *odpadki*) oraz heterogeniczne, pseudopoliczalne klasy „materiałów” i rzeczy podzielnych (np.: *narkotyki*, *warzywa* itp.).

„Wyrazistość psychologiczna jest skorelowana ze strukturą pojęciową w tym sensie, że kategorie kolektywne i funkcjonalne są generalnie biorąc niezbyt wyraziste. Jednak wyrazistość psychologiczna nie stanowi wystarczającej wskazówki co do struktury pojęciowej, ponieważ superkategorie taksonomiczne także mogą być relatywnie mało wyraziste. Implikacja działa jednokierunkowo. Zdolność do uzyskiwania wysokiej wyrazistości psychologicznej zależy od struktury semantycznej pojęcia: pojęciom nietaksonomicznym brak tej zdolności, którą wykazują pojęcia kontrastujące z nimi. To, czy dane pojęcie o odpowiedniej strukturze semantycznej uzyska czy też nie wysoką wyrazistość psychologiczną, zależy od jego roli kulturowej. Jak sugerował Dougherty, <<najbardziej wyraziste lub podstawowe kategorie w klasyfikacji biologicznej [...] zależą od indywidualnego lub kulturowego poziomu interakcji związanej z daną dziedziną>><sup>92</sup>.

Kategorie prototypowe mają zastosowanie w badaniu ekwiwalencji semantycznej zarówno między językami, jak i w nich samych. Wiadomo, że dokładne synonimy wewnątrzjęzykowe, czyli jednostki leksykalne o tym samym znaczeniu, wzajemnie wymienne we wszystkich kontekstach, są niezwykle rzadkie.

---

<sup>90</sup> Wierzbicka 1999, s. 89.

<sup>91</sup> Ibidem s. 90-91.

<sup>92</sup> Ibidem s. 108.

Geeraerts<sup>93</sup> badał czasowniki *vernieten* i *vernietigen* (zniszczyć coś, unieścić) w dziewiętnastowiecznym języku niemieckim. Obydwa odnoszą się do tego samego zakresu sytuacji i wykazują te same ograniczenia selekcyjne nawet w tekstach jednego i tego samego autora. Czy były zatem pełnymi synonimami? Geeraerts twierdzi, że nie. Różnice ujawniły się przy porównaniu ich różnych sensów, *vernietigen* stosowano głównie w sensie abstrakcyjnym, natomiast *vernieten* odnosił się przeważnie do aktów fizycznego niszczenia. Przykład ten wskazuje, że chociaż dwa wyrazy mogą być ekwiwalentne dystrybucyjnie i referencyjnie, to jednak mogą być związane z odrębnymi prototypami.

Taki stan rzeczy może zaistnieć w odniesieniu do „tego samego” słowa stosowanego przez użytkowników różnych odmian języka. Willett Kempton<sup>94</sup> przeprowadził eksperyment, w ramach którego zgromadził pięćdziesiąt rodzajów obuwia i prosił ankietowanych o ich poklasyfikowanie jako *boots*, *shoes*, *slippers* itd. Stwierdził, że teksascy i brytyjscy respondenci wykazali znaczną zgodność co do tego, które objekty są *boots*, a zatem zakresy odniesienia *boots* były porównywalne w obu grupach. Jednakże te dwie grupy respondentów wydawały się operować widocznie odmiennymi prototypami. Dla Teksaszczyków prototypowe *boots* sięgały wysoko powyżej kostki, jak buty kowbojskie; dla Brytyjczyków prototypem były kamasze, które nie sięgały powyżej kostki. Do jakiego stopnia zjawisko to jest charakterystyczne dla szerokiego zakresu obiektów obejmujących wszystkie dialekty (angielskie), czy nawet wówczas, gdy użytkownicy tego samego dialektu również posługują się różnymi prototypami – jest to problem zasługujący na dalsze badania. Warto przy tym zwrócić uwagę, że dla Langackera jest oczywiste, że nie ma dwóch użytkowników języka o dokładnie tym samym systemie językowym. Przypuszczalnie, w codziennej komunikacji różnice w centrach conceptualnych zachodzą w niezauważalny sposób, zwłaszcza jeśli odnośne jednostki leksykalne można stosować do identycznych zakresów odniesienia. Jednakże otwarte pozostaje pytanie, do jakiego stopnia różnice conceptualne mogą niekiedy utrudniać komunikację, a nawet prowadzić do poważnych jej zakłóceń. Niezależnie od zróżnicowania reprezentacji kategorii wśród użytkowników danego języka czy dialektu nie ulega wątpliwości, że reprezentacje wielu kategorii zmieniają się radykalnie z upływem czasu. Prototypowe rowery i samochody sprzed pięćdziesięciu lat obecnie stanowią marginesowe przykłady swoich kategorii. Być może, pojęcia bardziej abstrakcyjne, takie jak na przykład miłość czy piękno, podlegały jeszcze bardziej drastycznym zmianom.

Tabela 2 podsumowuje wyniki zastosowania badań Rosch nad kategorią MEBLE (niem. MÖBEL) w odniesieniu do niemieckojęzycznych respondentów proszonych o wskazanie stopnia przynależności do kategorii MEBLE niemieckich nazw pięćdziesięciu pięciu przedmiotów gospodarstwa domowego. Dirven, który przeprowadził ten eksperyment, sugeruje<sup>95</sup> że

<sup>93</sup> Geeraerts 1988. Cyt. za: Taylor 1991, s. 56.

<sup>94</sup> Kempton 1981. Cyt. za: Taylor 1991, s. 56-58.

<sup>95</sup> Cyt. za: Taylor 1991, s. 56.



Tabela 2. Reprezentatywność pięćdziesięciu pięciu okazów kategorii MEBLE

| Okaz                   | Ranga | Ocena trafności | Okaz                 | Ranga | Ocena trafności |
|------------------------|-------|-----------------|----------------------|-------|-----------------|
| łóżko                  | 1,5   | 1,00            | barek                | 29    | 3,70            |
| stół                   | 1,5   | 1,00            | łada                 | 30    | 3,73            |
| sofa                   | 3     | 1,13            | lustro               | 31    | 4,20            |
| szafka na<br>naczynia  | 5     | 1,20            | szuflada             | 32    | 4,38            |
| biurko                 | 5     | 1,20            | kredens              | 33    | 4,47            |
| szafa wnękowa          | 5     | 1,20            | lampa                | 34    | 4,79            |
| krzesło                | 5     | 1,20            | pianino              | 35    | 5,07            |
| fotel                  | 8     | 1,40            | zegar                | 36    | 5,20            |
| dwuosobowy<br>komoda   | 9     | 1,43            | kuchnia              | 37    | 5,33            |
| regał<br>książkowy     | 10    | 1,47            | kosz na śmieci       | 38    | 5,42            |
| nocna szafka           | 11    | 1,53            | lodówka              | 39    | 5,73            |
| stolik kawowy          | 12,5  | 1,67            | chodnik              | 40    | 5,80            |
| kanapa                 | 12,5  | 1,67            | obraz                | 41    | 5,93            |
| szafka na<br>porcelanę | 14    | 1,70            | klęcznik             | 42,5  | 6,00            |
| gablotka               | 16    | 1,73            | kotary               | 42,5  | 6,00            |
| materac na<br>stelażu  | 16    | 1,73            | telewizor            | 45    | 6,07            |
| sekretarzyk            | 16    | 1,73            | kołyska              | 45    | 6,07            |
| bufet                  | 18    | 1,80            | podnózek             | 45    | 6,07            |
| tapczan                | 19    | 1,93            | hi-fi                | 47    | 6,33            |
| półka                  | 20,5  | 2,00            | wazon                | 48    | 6,40            |
| fotel-bujak            | 20,5  | 2,00            | radio                | 49,5  | 6,47            |
| ława                   | 22    | 2,10            | maszyna do<br>szycia | 49,5  | 6,47            |
| kozetka                | 23    | 2,27            | wentylator           | 51    | 6,54            |
| taboret                | 23    | 2,27            | stereo               | 52    | 6,64            |
| serwantka              | 23    | 2,27            | poduszka             | 54    | 6,80            |
| otomana                | 26    | 2,43            | telefon              | 54    | 6,80            |
| pomocnik<br>(stolik)   | 27    | 2,80            | popielniczka         | 54    | 6,80            |
| stojak na gazety       | 28    | 3,40            |                      |       |                 |

Tabla 2a. Goodness-of-example ratings for fifty-five members of the category MÖBEL (FURNITURE)<sup>96</sup>

| Member           | Rank | Specific score | Member         | Rank | Specific score |
|------------------|------|----------------|----------------|------|----------------|
| bed              | 1,5  | 1,00           | bar            | 29   | 3,70           |
| table            | 1,5  | 1,00           | counter        | 30   | 3,73           |
| sofa             | 3    | 1,13           | mirror         | 31   | 4,20           |
| cupboard         | 5    | 1,20           | drawers        | 32   | 4,38           |
| desk             | 5    | 1,20           | dresser        | 33   | 4,47           |
| closet           | 5    | 1,20           | lamp           | 34   | 4,79           |
| chair            | 5    | 1,20           | piano          | 35   | 5,07           |
| love seat        | 8    | 1,40           | clock          | 36   | 5,20           |
| chest of drawers | 9    | 1,43           | stove          | 37   | 5,33           |
| bookcase         | 10   | 1,47           | wastebasket    | 38   | 5,42           |
| night table      | 11   | 1,53           | refrigerator   | 39   | 5,73           |
| cofee table      | 12,5 | 1,67           | rug            | 40   | 5,80           |
| couch            | 12,5 | 1,67           | picture        | 41   | 5,93           |
| china closet     | 14   | 1,70           | hassok         | 42,5 | 6,00           |
| cabinet          | 16   | 1,73           | drapes         | 42,5 | 6,00           |
| divan            | 16   | 1,73           | TV             | 45   | 6,07           |
| bureau           | 16   | 1,73           | rocker         | 45   | 6,07           |
| buffet           | 18   | 1,80           | footstool      | 45   | 6,07           |
| davenport        | 19   | 1,93           | hi-fi          | 47   | 6,33           |
| shelf            | 20,5 | 2,00           | vase           | 48   | 6,40           |
| rocking-chair    | 20,5 | 2,00           | radio          | 49,5 | 6,47           |
| bench            | 22   | 2,10           | sewing machine | 49,5 | 6,47           |
| lounge           | 23   | 2,27           | fan            | 51   | 6,54           |
| stool            | 23   | 2,27           | stereo         | 52   | 6,64           |
| vanity           | 23   | 2,27           | pillow         | 54   | 6,80           |
| ottoman          | 26   | 2,43           | telephone      | 54   | 6,80           |
| end table        | 27   | 2,80           | ashtray        | 54   | 6,80           |
| magazine rack    | 28   | 3,40           |                |      |                |

<sup>96</sup> Taylor 1991, s. 57

różnice między tabelami 1 i 2 świadczą, że niemieccy studenci prawdopodobnie oceniali przynależność do kategorii na podstawie umeblowania typowego pokoju studenta, podczas gdy respondenci amerykańscy mogli przywołać na myśl pokój dzienny w domu ich rodziców. Jeśli mamy do czynienia z takim przypadkiem, to MÖBEL i FURNITURE nie mogą być uznane za dokładne ekwiwalenty przekładowe, przynajmniej przy testowanej próbie, nawet jeśli wydawałoby się, że zakresy odniesienia (referencji) tych dwóch słów są zbieżne.

Kończąc, warto przypomnieć, że system kategorii może skutecznie funkcjonować wówczas, gdy nie zmienia się drastycznie za każdym razem, gdy pojawiają się nowe informacje. Równocześnie powinien być wystarczająco elastyczny, aby móc się dostosowywać do zmian otoczenia. Z pewnego punktu widzenia kategoriami najdoskonalszymi, opartymi na całkowitej korelacji cech wszystkich okazów kategorii, są kategorie klasyczne. Jednakże ich sztywność czyni je niewystarczającymi z punktu widzenia różnych potrzeb ludzkiego poznania. W pewnych sytuacjach kategorie prototypowe, w istotnym stopniu spełniające połączone warunki stabilności i elastyczności<sup>97</sup>, mogą skutecznie zastąpić kategorie Arystotelesowskie. Centralne okazy kategorii prototypowej mają dużą liczbę wspólnych atrybutów i pod tym względem centrum kategorii prototypowej zbliża się do ideału klasycznej kategorii. Równocześnie kategorie prototypowe umożliwiają przynależność obiektom, które mają mniej czy niewiele atrybutów wspólnych z okazami centralnymi, ale w świadomości użytkowników języka są z nimi kojarzone.

Dalecy jesteśmy od sugerowania, że jedna z koncepcji kategoryzacji jest poprawna, a druga nie. Naszym zdaniem te dwa podejścia nie tworzą alternatywy. Przeciwnie, rozpatrywane łącznie mogą stanowić komplementarną postawę metodologiczną. Pogląd taki wyraziła także Wierzbicka, pisząc „w analizie semantycznej jest miejsce dla prototypów, ale jest również i dla inwariantów – jedno nie wyklucza drugiego”<sup>98</sup>.

Podobnie, doceniając liczne zalety Wittgensteinowskiej koncepcji znaczenia jako użycia, dalecy jesteśmy od negowania wartości rzetelnych definicji, tam, gdzie można je sformułować. W niektórych pracach z zakresu językoznawstwa, psychologii czy nauk kognitywnych daje się zauważyć tendencja do wyolbrzymiania roli prototypów i traktowania ich jako panaceum na wszystkie problemy. Trudności, jakie mogą się pojawić przy formułowaniu definicji, nie są wystarczającym powodem do zaniechania prób opracowywania adekwatnych definicji i zastępowania ich pseudodefinicjami opartymi na prototypach.

---

<sup>97</sup> Z jednej strony powstawanie niuansów pojęciowych świadczy o ich dynamicznej zdolności nadążania za zmieniającymi się warunkami i potrzebami ekspresji. Z drugiej strony, fakt, że pojęcia marginalne można przyłączyć do istniejących kategorii jako ich peryferyjne przypadki dowodzi, że kategorie wykazują skłonność do utrzymywania się jako jednostki holistyczne, a więc zachowując ogólną strukturę systemu kategoryalnego.

<sup>98</sup> Wierzbicka 1999, s. 27.

Kategorie prototypowe nie nakładając obowiązku zachowywania jednolitego kryterium doboru atrybutów, pozwalają na porządkowanie rzeczywistości na podstawie różnych wybranych atrybutów, na skupianie się na jednych a pomniejszanie innych, podczas gdy klasyfikacja czy kategoryzacja klasyczna dokonywana jest na podstawie wszystkich znanych cech według jednolitego kryterium. Wykorzystanie koncepcji kategorii prototypowych pozwoliłoby projektantom języków informacyjno-wyszukiwawczych zorganizować pola znaczeniowe jiw wokół wyrazistych przykładów kategorii, dając jednocześnie możliwość kilkakrotnego przyporządkowania przypadków mniej wyraźnych. Owa mniejsza wyrazistość nie musi oznaczać indywidualnej kategoryzacji użytkownika. Równie dobrze może być prezentowana przez autorów dokumentów, którzy modelując obraz rzeczywistości świadomie eksponują pewne własności, atrybuty itp., pomniejszając czy wręcz eliminując inne. W językach opartych na podziale dziedzinowym czy koniunkcji cech koniecznych i wystarczających idiosynkrazje obrazów rzeczywistości przedstawionych w dokumentach muszą być traktowane jako nierелеwantne. Wykorzystanie kategorii prototypowych pozwoliłoby także na wyeksponowanie tych elementów znaczenia, które są uwarunkowane kulturowo i których ekspozycja w innej sytuacji byłaby trudna czy wręcz niemożliwa. Nie można bowiem zapominać, że dokumenty gromadzone w bibliotekach nie zawsze odwzorowują kategorie naukowe. Podobne pytania kierowane do bibliotek operują zarówno klasycznymi kategoriami naukowymi, jak i kategoriami ogólnej wiedzy o świecie i języka potocznego. Sytuacje te wymagałyby zatem, jeśli nie uwzględnienia w projektach jiw, to przynajmniej świadomości ograniczeń języka operującego tylko jednym podejściem.



## KATEGORYZACJA W TEORII I PRAKTYCE JĘZYKÓW INFORMACYJNO-WYSZUKIWAWCZYCH

Kategoryzacja w teorii i praktyce języków informacyjno-wyszukiwawczych bywa różnie rozumiana<sup>1</sup>. Najczęściej kategoryzacją nazywa się albo realizowany na etapie projektowania lub modyfikowania jiw proces wyodrębniania obiektów i ich zbiorów traktowanych z pewnego punktu widzenia jako jednorodne, czyli proces tworzenia systemu kategorii w języku, albo rezultat tak rozumianej kategoryzacji, czyli system kategorii. Kategoryzacja jest również rozumiana jako kategoryzowanie, czyli proces przypisywania pewnych obiektów (informacji zawartych w dokumentach i informacji o dokumentach) do kategorii już istniejących w języku informacyjno-wyszukiwawczym.

Literatura dotycząca problematyki kategoryzacji w pierwszym i trzecim ze znaczeń jest dość uboga. Najwięcej publikacji dotyczy kategoryzacji rozumianej jako system kategorii. Proces dochodzenia do systemu kategorii i, ogólniej, pewnej struktury semantyczno-leksykalnej języka informacyjno-wyszukiwawczego rzadko jest przedmiotem publikacji. Również rzadko autorzy zajmują się relacjami pomiędzy obrazem świata w jiw a obrazem utrwalonym w dokumentach i obrazem świata samych dokumentów. Problemy te wymagają wnikliwych i kompleksowych badań teoretycznych i empirycznych, i to nie tylko w aspekcie kategoryzacji.

Często autorzy publikacji stosują termin *kategoryzacja* wymiennie<sup>2</sup> z *klasyfikacja*<sup>3</sup>. Precyzyjne rozdzielenie znaczeń tych dwóch terminów i, co za tym idzie, zakresów ich stosowania nie jest obecnie możliwe, co widać szczególnie

---

<sup>1</sup> Wieloźnaczność *kategoryzacji* charakterystyczna jest nie tylko dla nauki o informacji, o czym pisaliśmy w rozdz. drugim.

<sup>2</sup> Warto jednak zauważyć, że zwykle wymiennosc ta jest jednokierunkowa, tj. *klasyfikowanie* i *klasyfikacja* są rozumiane jako synonimy *kategoryzowania* i *kategorii*, ale nie odwrotnie. Wyjątkiem jest objaśnienie *kategoryzacji* (tekstów literackich) zawarte w *International encyclopedia of information and library science* (International 1997a, s. 50), gdzie *kategoryzacja* potraktowana jest jako synonim *klasyfikacji* („In the USA the term classification tends to be used to mean categorization”).

<sup>3</sup> Dodatkową komplikację stanowi wieloźnaczność samego terminu *klasyfikacja*, który w nauce o informacji najczęściej używany jest w czterech znaczeniach: 1) dowolny podział

wyraźnie w przypadku klasycznego ujęcia kategoryzacji. Klasyfikacja i kategoryzacja mają wiele wspólnych cech i problemów, zwłaszcza jeśli się weźmie pod uwagę wymóg naturalności podziału logicznego. Przypomnimy, że klasyfikacja to wielostopniowy podział logiczny, na który dodatkowo mogą być nałożone pewne warunki o charakterze pozaformalnym (patrz: s. 119). W przeciwieństwie do klasyfikacji, kategoryzacja, oczywiście poza ujęciem klasycznym, nie implikuje ani tego, że przynależność do kategorii jest wyznaczana przez postrzeganie cech koniecznych i wystarczających, ani tego, by zawieranie się w jednej kategorii wykluczało przynależność do innej. Uznanie podobieństw między skądinąd niepodobnymi jednostkami i wynikająca stąd identyfikacja kategorii pozwala odkrywać ład w rzeczywistości. Można zatem nieco przewrotnie stwierdzić, że klasyfikacja opiera się na postrzeganiu podobieństw i, co za tym idzie, jest manifestacją jednorodności, podczas gdy kategoryzacja ma za podstawę postrzeganie różnic i wspiera różnorodność.

Języki informacyjno-wyszukiwawcze typu klasyfikacji w różnym stopniu spełniają warunki bycia podziałem logicznym. Najczęściej spotykanymi w nich odstępstwami są:

– jednoczesny podział klasy macierzystej według kilku zasad, na przykład jeśli klasę **Gazy** podzielono by na klasy pochodne: **Gazy trujące**, **Gazy nietrujące**, **Gazy palne** i **Gazy niepalne**, mielibyśmy do czynienia z przykładem zastosowania dwóch kryteriów podziału, „bycia trującym” i „bycia palnym”, na tym samym stopniu klasyfikacji, czyli z brakiem jedności zasady podziału;

– nierozłączność zakresów klas pochodnych, będąca konsekwencją braku jedności zasady podziału;

– nieprzestrzeganie zasady zupełności klasyfikacji, czyli tego, że suma zakresów klas pochodnych powinna być równa zakresowi klasy macierzystej;

– traktowanie klasy macierzystej jako sumy klas pochodnych, a nie jako klasy o charakterze ogólniejszym w stosunku do klas pochodnych – w takim przypadku w ogóle nie można mówić o istnieniu zasady podziału klasy macierzystej; ilustracją może być podział klasy macierzystej **Organizacja i zarządzanie przedsiębiorstw. Organizacja łączności i transportu. Przemysł wydawniczy** na klasy: **Organizacja biura. Biurowość; Sposoby pisania i powielania pism. Tłumaczenia; Skróty. Stenografia; Organizacja łączności; Przemysł wydawniczy. Drukarstwo. Księgarstwo; Organizacja transportu. Poczta; Rachunkowość. Księgowość; Organizacja przedsiębiorstw (zakładów); Propaganda gospodarcza. Reklama. Sprawozdawczość**<sup>4</sup>;

---

logiczny; 2) podział logiczny rozgałęziony; 3) język informacyjno-wyszukiwawczy, którego paradygmatykę stanowią relacje hierarchiczne, najczęściej odwzorowane w strukturze jego wyrażień; 4) stosowanie klasyfikacji w znaczeniu 3) do tworzenia charakterystyk i instrukcji wyszukiwawczych (inaczej klasyfikowanie). Tylko pewne jiw typu klasyfikacji spełniają (w różnym stopniu) warunki bycia podziałem logicznym lub podziałem logicznym rozgałęzionym. Będzie o tym mowa w rozdziale *Kategorie w językach informacyjno-wyszukiwawczych*.

<sup>4</sup> Przykład przejęto z: Dobrowolski 1956, s. 16.

– niezgodność zakresów klasy macierzystej i klas pochodnych – np. jeśli klasę macierzystą **Rodzina** podzielono by na klasy pochodne: **Dziecko**, **Macierzyństwo**, **Małżeństwo**, **Monogamia**, **Poligamia**, **Poliandria**, **Władza rodzicielska**, to mielibyśmy do czynienia z przykładem niezgodności zakresów klasy macierzystej i klas pochodnych przy jednoczesnym braku jedności zasady podziału.

U podstaw zarówno klasyfikacji, jak i kategoryzacji leży pojęcie porządku. Porządkowanie w szerokim sensie oznacza każde układanie czegoś według określonego planu. Relacją całkowicie porządkującą jest stosunek jednocześnie spójny (zachodzi między dwoma elementami porządkowanego zbioru), przechodni (zachodząc między elementami  $x$  i  $y$  oraz między  $y$  i  $z$ , zachodzi też zawsze między  $x$  i  $z$ ) i antysymetryczny (zachodząc między  $x$  i  $y$ , nie zachodzi nigdy między  $y$  i  $x$ ). Porządkiem jest na przykład relacja następstwa kolejnych wyrazów w zdaniu. Zabiegiem ogólnoporzadkującym jest grupowanie elementów zbioru, czyli tworzenie ich zespołów. Jeśli przedmioty w każdej grupie będą pod określonym względem jednorodne, to otrzymamy klasy; jeśli zaś nie, wówczas możliwa jest typologia, czyli zabieg (lub rezultat takiego zabiegu) polegający na wyróżnieniu (w obrębie danego zbioru) przedmiotów wzorcowych (realnych lub idealnych, ale posiadających cechy zdecydowane i powtarzające się u innych – typów) i grupowaniu reszty przedmiotów wokół wyróżnionych wzorców.

Na zespołach przedmiotów, a właściwie na zakresach nazw i pojęć, możemy dokonywać podziału logicznego. W wyniku spiętrzenia podziałów logicznych powstaje klasyfikacja będąca podziałem wielostopniowym. Gdy klasyfikację stosuje się do dziedziny jednorodnych przedmiotów empirycznych, otrzymujemy systematykę. Jeśli klasyfikowanymi obiektami są zwierzęta i/lub rośliny, to systematyka taka bywa też nazywana taksonomią. Formalnymi warunkami poprawności podziału klasyfikacyjnego są zupełność (zakresy członów podziału<sup>5</sup> muszą być adekwatne i wyczerpujące w stosunku do zakresu całości, tzn. suma zakresów członów podziału musi się równać zakresowi pojęcia dzielonego) i rozłączność (zakresy członów nie mogą się krzyżować)<sup>6</sup>. Wśród warunków pozaformalnych zwraca się najczęściej uwagę na to, aby: 1) zakresy

---

<sup>5</sup> Występujące w podziale logicznym pojęcie nadrzędne nazywa się *pojęciem dzielonym*, a pojęcia podrzędne – *członami podziału*; cecha, według której dokonuje się podziału, to *zasada* lub *podstawa podziału*.

<sup>6</sup> Najłatwiej zachować adekwatność i rozłączność w podziałach dwuczłonowych (dychotomicznych), czyli przeprowadzanych według cech sprzecznych. Takim podziałem jest na przykład podział mieszkańców miasta  $X$  w Polsce na pełnoletnich i niepełnoletnich. Trudniej zachować adekwatność i rozłączność, gdy człony podziału powstają w wyniku dołączania do treści pojęcia dzielonego różnych odmian cechy będącej zasadą (podstawą) podziału, przy dodatkowym spełnieniu następujących warunków: 1) uwzględnione zostały wszystkie odmiany cechy stanowiącej podstawę podziału; 2) cecha będąca podstawą podziału przysługuje wszystkim dezygnatom pojęcia dzielonego; 3) żaden przedmiot leżący w zakresie pojęcia dzielonego nie posiada dwóch (lub więcej) odmian cechy będącej zasadą podziału. Według takiego schematu dzielimy na przykład ludzi na trzy podstawowe rasy – białą, czarną i żółtą.



klas posiadały dostateczną ostrość; 2) zakresy cząstkowe były tworzone w sposób naturalny, tak, żeby elementy tej samej klasy posiadały jak najwięcej cech wspólnych, były bardziej do siebie podobne (z interesującego nas punktu widzenia) niż elementy należące do różnych klas; 3) zasada podziału została dostosowana do zadań; 4) klasyfikacja była ciągła (stopniowo, bez skoków, przechodziła do bardziej cząstkowych zakresów).

Klasyfikację (wielostopniowy podział logiczny) dobrze ilustruje taksonomia (systematyka biologiczna), czyli klasyfikacja jednorodnych przedmiotów biologicznych, służąca opisywaniu gatunków, ich nazywaniu i włączaniu w układ systematyczny zwierząt i roślin. Cel taksonomii jest uzyskiwany poprzez ustanowienie według określonych i ogólnie akceptowanych zasad i praw hierarchicznego układu wzajemnie rozłącznych klas. Na podstawie tych praw każda jednostka jest oceniana, określone są jej cechy zasadnicze czy definiujące, a następnie jest przyporządkowywana do jednej klasy w ramach hierarchii. Grupując, na przykład, podobne do siebie rośliny w niewielkie stosunkowo klasy, te klasy w klasy wyższego rzędu itd. otrzymujemy coraz mniejszą liczbę klas obejmujących coraz większą liczbę roślin, aż wreszcie dwie wielkie klasy roślin jedno- i wielokomórkowych łączące się we wspólną klasę-uniwersum ROŚLINY. Każdej klasie przypisana jest unikatowa nazwa (w przypadku taksonomii biologicznej – łacińska), implikująca zbiór cech dystynktywnych, które odróżniają tę klasę od innych klas w hierarchii. Klasy mają wyraźne i dobrze określone granice: dana jednostka musi odwzorowywać pełny zakres zasadniczych cech, które są specyficzne dla określonej klasy i jedynie te obiekty, które posiadają ten zbiór cech dystynktywnych są identyfikowane przez unikatową nazwę przypisaną danej klasie. Z kolei nazwa, która jest uniwersalnie stosowana do identyfikowania okazów specyficznej klasy, służy jako znacznik umożliwiający dostęp do informacji o jednostkach w ramach tej klasy. Klasyfikacja taksonomiczna w pewnym sensie stanowi streszczenie klasycznej teorii kategoryzacji: obiekt biologiczny jest pełnym i równorzędnym członem specyficznej klasy, jeśli odznacza się cechami zasadniczymi lub dystynktywnymi, które są niezbędne i wystarczające do uznania przynależności do kategorii. Klasyfikacja, podobnie jak klasyczna teoria kategorii, opiera się na założeniu, że bycie okazem określonej klasy (ekstensja) implikuje posiadanie cechy (cech) zasadniczej lub dystynktywnej (intensja) charakteryzującej tę klasę jako niepowtarzalną całość.

Kategoryzacja spełnia fundamentalną funkcję w procesach poznania. Poprzez dostrzeganie i uznawanie podobieństw między potencjalnie odmiennymi przedmiotami człowiek uzyskuje możliwość rozszerzania uogólnień utworzonych na podstawie poprzednich doświadczeń. Bez zdolności rozpoznawania podobieństw, istniejących między skądinąd niepodobnymi obiektami, byłibyśmy przytłoczeni złożonością rzeczywistości. Kategoryzacja jest środkiem jej upraszczania, redukcji obciążenia pamięci, pomaga w efektywnym gromadzeniu i wyszukiwaniu informacji. Stwierdzenie to odnosi się zarówno do sytuacji ogólnych, codziennych, jak i sytuacji informacyjnych.

Podobnie jak klasy w klasyfikacji, kategorie nie istnieją w izolacji, lecz często są łączone relacjami inkluzji, obejmującymi szczególnie przypadki w ramach struktury szerszych czy bardziej ogólnych kategorii nadrzędnych.

W krytycznym omówieniu pracy Edwarda Smitha i Douglasa Medina<sup>7</sup> filozof Georges Rey<sup>8</sup> wyróżnił cztery pierwotne funkcje realizowane przez kategorie: funkcję stabilności, zapewniającą spójność reprezentacji mentalnych zarówno u poszczególnych jednostek, jak i wśród nich; funkcję językową, wyposażającą wyrażenia języka w znaczenia poprzez relacje synonimii, antynomii, implikację semantyczną oraz przekład; funkcję metafizyczną, która definiuje i określa ważność identyfikacji systematyzacyjnych, co znaczy, że dana identyfikacja jest prawdziwa metafizycznie, jeśli jest prawdziwa we wszelkich możliwych światach; i wreszcie funkcję epistemologiczną, zapewniającą możliwość epistemologicznej identyfikacji obiektu w ramach kategorii – to znaczy jej spójność z poszerzoną wiedzą jednostki o świecie.

W reakcji na propozycję Reya, Medina i Smith<sup>9</sup> zaproponowali swój własny zestaw czterech funkcji, czyli ról pełnionych przez pojęcia, przy czym termin pojęcie ma się odnosić do mentalnych reprezentacji kategorii prostej. Prosta kategoryzacja jest opisywana jako proces mający na celu określenie, czy dany obiekt jest, czy też nie jest przypadkiem określonej kategorii, takiej jak np. PIES (ang. DOG). Kompleksowa kategoryzacja wyznacza przynależność do kategorii złożonej pojęciowo, takiej jak np. ZŁY PIES (ang. MEAN DOG). Znaczenie językowe, jedyna funkcja przypisana pojęciom przez Medina i Smitha, która wykazuje podobieństwo do funkcji kategorii opisanych przez Reya, wyjaśnia relacje semantyczne między słowami, tak jak w odniesieniu do synonimii między *pies* i *canine*, czy antynomii między *pies* i *kot*. Ostatnia z ról opisanych przez Medina i Smitha, składniki stanów poznawczych, reprezentuje coś, co określają jako „krytyczny składnik” przekonania: tak jak składniki stanów poznawczych, „pojęcia są tym, co dostarcza kognitywnego wyjaśnienia kompleksowego myślenia i zachowania”<sup>10</sup>. Zatem proste kategorie ZŁY, PIES i GRYŹĆ współdziałają ze sobą tworząc konceptualne przekonanie, że *złe psy gryzą* (ang. *mean dogs bite*).

Jakkolwiek Medina i Smith wskazują Reya jako pierwotne źródło wyróżnionych przez nich funkcji poznawczych przypisywanych kategoriom, rozbieżność między tymi dwoma podejściami jest dość duża. Oczywiście oparcie się przez Medina i Smitha na epistemologicznej interpretacji kategorii (pojmowania rzeczywistości przez jednostkę) aż do wykluczenia aspektów metafizycznych (fundamentalny charakter rzeczywistości rozprzestrzeniający się na wszystkie możliwe światy) ujawnia konflikt, który przenika nowsze badania nad procesem kategoryzacji i naturą kategorii. Klasyczna teoria kategorii zakłada, że ekstensja równa jest intensji – czyli, że bycie okazem specy-

<sup>7</sup> Smith 1981.

<sup>8</sup> Rey 1983. Cyt. za: Jacob 1991, s. 79-80.

<sup>9</sup> Medina 1984. Cyt. za: Jacob 1991.

<sup>10</sup> Ibidem s. 80.

ficznej kategorii implikuje posiadanie zasadniczej i dystynktywnej cechy, która przebiega przez wszystkie możliwe klasy w jednym łańcuchu hierarchicznym. Jednakże badania wykazały, że klasyczna teoria nie dysponuje wystarczającą siłą eksplikacji, by wyjaśnić niestabilność przynależności do kategorii, odzwierciedlającą różnorodność i elastyczność, z jaką ludzie tworzą kognitywne reprezentacje rzeczywistości.

Analityczna organizacja systematyk naukowych oraz sztuczna struktura konceptualna nałożona na układ wiedzy, a właściwie piśmiennictwa, przez schematy klasyfikacji biblioteczno-bibliograficznych mają wiele wspólnego z klasyczną teorią kategoryzacji. Obie są sztywno ograniczone przez logiczne, lecz tym nie mniej arbitralne, uporządkowanie rzeczywistości. Jednakże ludzie rzadko podlegają tego typu ograniczeniom w kognitywnym porządkowaniu rzeczywistości; ograniczenia wypływają raczej z zasad komunikowania się. Wymagania nałożone przez potrzebę komunikowania zmuszają do zastępowania własnych reprezentacji świata bardziej stabilnymi strukturami kategorii, wspólnymi dla różnych osób. Fundamentalne znaczenie ma zrozumienie, w jaki sposób jednostki stosują kategorie, by ustrukturalizować ich własną wiedzę. Równie ważne jest zrozumienie, jak kategorie mogą być ustrukturalizowane, czy restrukturyzowane przez zamiar komunikowania. „Prawdopodobieństwo funkcjonowania kategorii zgodnie z klasyczną teorią na jednym poziomie przy ilustrowaniu aspektów niestabilności, stopniowalnej struktury i ich rozmytych granic na innym poziomie, powinno zachęcić badaczy do odłożenia, na pewien czas, zainteresowania strukturą kategorii i do skoncentrowania uwagi na generowaniu kategorii kognitywnych oraz roli (ról) pełnionych przez te idiosynkratyczne<sup>11</sup> kategorie w procesie poznania i komunikacji. Uznanie, że to potrzeba komunikowania może zmuszać jednostki do restrukturyzowania ich mentalnych reprezentacji zgodnie ze zbiorem zasadniczych cech lub „wspólnych wartości czy wymiarów” wskazuje drogę do poszerzonego rozumienia tego, w jaki sposób idiosynkratyczne, osobiste kategorie wytwarzane przez jednostkę współdziałają z uniwersalnymi, analitycznymi kategoriami stosowanymi w klasyfikacji. Ważne jest tu nie pytanie: „Jaka jest struktura kategorii”, lecz „Czy przy ograniczeniach nakładanych przez potrzebę komunikacji, jednostka restrukturyzuje zinternalizowane (wewnętrzne) reprezentacje rzeczywistości – te idiosynkratyczne kategorie kognitywne – by odzwierciedlić ustalony i arbitralny układ wiedzy? A jeśli tak, to w jaki sposób? Odpowiedzi na to złożone pytanie powinny poszerzyć rozumienie funkcjonalnych związków istniejących między klasyfikacją, poznaniem i tworzeniem kategorii. Jest bardzo prawdopodobne, że mogą one wskazać kierunek poprawy dostępu do materiałów zorganizowanych zgodnie z obecnymi schematami klasyfikacyjnymi”<sup>12</sup>.

---

<sup>11</sup> Idiosynkratyczne – własne, osobiste, nie wynikające z reguł ogólnych czy powszechnej akceptacji. [Przypis JW.]

<sup>12</sup> Jacob 1991, s. 81-82.

## KATEGORIE W JĘZYKACH INFORMACYJNO-WYSZUKIWAWCZYCH

Kategoryzacja jest jednym, obok relacji paradygmatycznych i podziałów dziedzinowych, z podstawowych środków organizacji systemu semantyczno-leksykalnego języków informacyjno-wyszukiwawczych. Niektórzy autorzy, jak na przykład Elżbieta Artowicz<sup>13</sup>, uważają, że relacje paradygmatyczne są wykładnikami językowych kryteriów relewancji. Naszym zdaniem prawdziwość tego poglądu zależy od sposobu interpretacji relacji mereologicznej i relacji hierarchii tematycznej, na co zwróciła uwagę sama Artowicz<sup>14</sup>.

W literaturze polskiej problematyka relacji paradygmatycznych wydaje się być wyjątkowo obszernie i dokładnie zbadana<sup>15</sup>, choć z różnymi konsekwencjami dla praktyki. Relacje paradygmatyczne są systemowymi relacjami semantycznymi. Do ich identyfikacji powinna wystarczyć tylko i wyłącznie znajomość struktury systemu językowego. W teorii jiw terminem relacje paradygmatyczne obejmuje się szerszą klasę takich związków semantycznych pomiędzy jednostkami leksykalnymi jiw, które są *explicite* zaznaczone w słownikach tych języków i odzwierciedlają albo relacje językowe, albo stałe zależności między elementami rzeczywistości pozajęzykowej. Najczęściej wyróżnia się trzy<sup>16</sup> podstawowe rodzaje relacji paradygmatycznych:

– synonimię, będącą językową relacją równoważności, wyodrębnianą w tych językach informacyjno-wyszukiwawczych, w których nie wyznacza się form preferowanych (przyjętych), a wyszukiwanie musi uwzględniać możliwość reprezentowania znaczenia przez wiele (więcej niż jedną) informacyjnie i wyszukiwawczo ekwiwalentnych form;

– hierarchię, obejmującą, poza zwykłą hipero- i hiponimią, także inne charakterystyczne dla jiw związki o własności przechodniości i przeciwsymetryczności, tj. związek mereologiczny i hierarchii tematycznej;

– asocjacje (kojarzenia, skojarzenia), obejmujące zróżnicowaną klasę związków semantycznych ważnych ze względu na metainformacyjną i wyszukiwawczą funkcję jiw.

W teorii języków informacyjno-wyszukiwawczych przyjęto, że jedynie relacje hierarchiczne mogą pełnić funkcję organizującą (porządkującą) leksykę. Prowadzą one do utworzenia drzewa (struktura monodendrytowa jiw, zwana też monohierarchiczną lub jednozrębową) lub drzew klasyfikacyjnych (struktura polidendrytowa, zwana też polihierarchiczną lub wielozrębową)

---

<sup>13</sup> Artowicz 1997.

<sup>14</sup> Ibidem s. 93.

<sup>15</sup> Między innymi: Marciszewski 1973; Sosińska 1988; Ungurian 1976; Wereszczyńska-Cisło 1986a, 1984.

<sup>16</sup> Autorzy różnie się odnoszą do liczby relacji paradygmatycznych w jiw. Maksymalnie przyjmując się, że wszelkie relacje paradygmatyczne w jiw można sprowadzić do jednego z pięciu następujących związków: synonimii, hiperonimii, hiponimii, krzyżowania, wykluczania (antonomii, współrzędności, wykluczania mocnego).

odwzorowujących pewne semantyczne układy hierarchiczne jednostek leksykalnych. Struktura monodendrytowa charakteryzuje się tym, że na każdym stopniu podziału stosowana jest tylko jedna zasada podziału<sup>17</sup>, w efekcie czego wszystkie wyodrębnione klasy układają się w jedno drzewo (jeden zrąb, jeden układ hierarchiczny). W strukturze monodendrytowej nie ma możliwości odróżniania odmiennych kategorii słownictwa. Struktura polidendrytowa, w której na jednym stopniu podziału są stosowane co najmniej dwie zasady podziału, często opiera się na wcześniejszym podstawowym podziale kategoriowym wyrażań, wyznaczającym niezależne układy hierarchiczne leksyki jiw. Klasy pochodne, wyodrębnione w rezultacie zastosowania na jednym stopniu różnych zasad podziału, zwane są fasetami<sup>18</sup> (czasem także podkategoriami). Wyrażenia jiw wchodzące w skład jednej fasety mają wspólną cechę semantyczną, będącą implikacją zastosowanego kryterium wyodrębnienia fasety. W poszczególnych układach hierarchicznych struktury polidendrytowej istnieje możliwość monorelacyjnej organizacji drzew hierarchicznych.

W dotychczasowej teorii jiw ideałem był podział logiczny pola semantycznego jiw<sup>19</sup>. W praktyce jednak warunek ten udawało się spełnić tylko w ograniczonym zakresie. Obserwację tę radykalnie wyraził Eugeniusz Ścibor, pisząc, że termin klasyfikacja w odniesieniu do języków informacyjno-wyszukiwawczych należałoby widzieć w dwóch znaczeniach, z których szersze (znaczenie węższe to wielostopniowy podział logiczny) oznacza „wynik podziału obiektów (którymi w interesującym nas przypadku są dokumenty) na pewne klasy, przy czym nie zakłada się ani istnienia układu hierarchicznego (tj. wyróżnienia klas i podklas), ani podziału rozłącznego i adekwatnego; w tym znaczeniu klasyfikacją możemy nazwać wynik jakiegokolwiek podziału zbioru na podzbiory. (...) w znaczeniu szerszym klasyfikacją jest każdy język informacyjny, ponieważ każdy język informacyjny pozwala podzielić ogół dokumentów na pewne ich klasy. W drugim z podanych znaczeń (tj. w znaczeniu węższym) tylko niektóre języki informacyjne możemy nazwać klasyfikacjami”<sup>20</sup>.

Relacja hierarchii tematycznej zwana też relacją hierarchii dokumentacyjnej zachodzi nie między zakresami nazw, ale między zbiorami dokumentów scharakteryzowanych za pomocą tych nazw. Wiąże na przykład „wyrażenia

---

<sup>17</sup> Postulat jednej zasady podziału bywa niekiedy mylony z monorelacyjnością. W praktyce nie spotyka się języków informacyjno-wyszukiwawczych o monorelacyjnej strukturze pola semantycznego, czyli takiej, której wszystkie elementy byłyby powiązane relacją jednego typu.

<sup>18</sup> Termin faseta najczęściej jest używany w dwóch znaczeniach: 1) klasa leksykalna, której wszystkie elementy charakteryzują się wspólnym pierwiastkiem znaczeniowym, reprezentującym aspekt, cechę, właściwość itp., w jakiej kategoria jest rozpatrywana w danej fascie (wówczas faseta bywa też nazywana podkategorią); 2) zasada podziału, według której zostały zgrupowane wyrażenia klasyfikacji fasetowej.

<sup>19</sup> W rozdziale synonimicznie traktujemy terminy pole semantyczne (znaczeniowe) i pole leksykalne (wyrazowe, znakowe); ich znaczenia dokładniej będą objaśnione w rozdziale *Tworzenie kategorii w językach informacyjno-wyszukiwawczych*.

<sup>20</sup> Ścibor 1996, s. 32.

reprezentujące pewne dyscypliny wiedzy określane jako tematy szersze z wyrażeniami reprezentującymi obiekty lub problemy badawcze tych dyscyplin określane tematami węższymi wobec tamtych, lub też wyrażenia reprezentujące pewne obiekty badawcze wyznaczające szersze zakresowo tematy z wyrażeniami reprezentującymi pewne zagadnienia badania tych obiektów wyznaczające tematy zawierające się w tamtych, np. Semiotyka – Znaki, Fizjologia wysiłku – Próg anaerobowy, Psychologia – Osobowość, Ekonomika i finansowanie turystyki – Efektywność ekonomiczna turystyki, Prawo kościelne – Kary kościelne, Ptaki – Migracja ptaków, Zioła – Uprawa ziół, Nowotwory – Inwazyjność nowotworów, Teoria treningu – Obciążenie treningowe”<sup>21</sup>. Zdaniem Sosińskiej relacja ta nie ma jednoznacznego charakteru relacji językowej lub pozajęzykowej. Naszym zdaniem relacja hierarchii tematycznej jest związkiem pozajęzykowym. Ze względu na swój specyficznie porządkujący charakter jest wykorzystywana do organizacji mniejszych lub większych fragmentów pola leksykalnego jiw. W literaturze brak jest informacji o tym, jak w różnych językach informacyjno-wyszukiwawczych były ustalone związki hierarchii tematycznej. Teoretycznie, aby określić, czy w jakimś konkretnym przypadku zachodzi ta relacja, potrzebne byłyby badania dokumentów z danej dziedziny. Ustanawianie relacji hierarchii tematycznej powinno więc być oparte na kryterium piśmiennictwa (ang. *literary warrant*), tj. kryterium organizacji informacji na podstawie cech, atrybutów, własności itp. dokumentów opisanych w systemie i zarejestrowanego w nich obrazu rzeczywistości. Można jednak przypuszczać, że w praktyce nie zawsze tak się dzieje i że ustalenie członu nadrzędnego i członu podrzędnego w przypadku tego typu związku jest kwestią decyzji arbitralnej. „Idąc tropem relacji hierarchii tematycznej w ujęciu indeksowym można bezbłędnie dochodzić do potrzebnych dokumentów na jakiś temat nadrzędny, wychodząc od jakiegokolwiek tematu podrzędnego, na przykład od deskryptora «rodzina» do dokumentów charakteryzowanych deskryptorem «małżeństwo». Takiej niezawodności nie gwarantuje np. hierarchia zakresowa, bo kiedy termin «mysz» kieruje nas w teaurusie do terminu «ssak» jako nadrzędnego, wcale nie daje to gwarancji, że w każdym dokumencie dotyczącym myszy znajdują się informacje o ssakach jako takich; wszak badanie naukowe zatrzymuje się na pewnym zamierzonym stopniu ogólności i choć mówiąc o desygnatach jakiegokolwiek pojęcia mówi się zarazem o desygnatach każdego z pojęć zakresowo nadrzędnych, to nie jest to przecież mówienie na temat owych pojęć nadrzędnych: twierdząc cokolwiek na temat myszy mówi się tym samym o pewnych ssakach, ale nie znaczy to, że mówi się na temat ssaków; a przecież tego potrzebowałby użytkownik systemu informacyjnego poszukujący danych na temat ssaków i wychodzący, w poszukiwaniu tych danych, również od pojęć podrzędnych”<sup>22</sup>.

<sup>21</sup> Sosińska 1988, s. 157-158.

<sup>22</sup> Marciszewski 1977, s. 206-207.

Najbardziej zróżnicowaną grupę relacji stanowią relacje asocjacyjne. Należą doń związki typu: podobieństwo, zróżnicowanie, przyczyna – skutek, styczeń w czasie, styczeń w przestrzeni, obiekt – własność, obiekt – operacja, obiekt – składniki, metoda – instrument, proces – wykonawca i wiele innych. Badania eksperymentalne<sup>23</sup> wykazały, że zarówno liczba relacji asocjacyjnych, jak i poziom ich specyfikacji mają korzystny wpływ na efektywność wyszukiwania informacji w systemach informacyjno-wyszukiwawczych.

Niekiedy pojęcie kategoryzacji w językach informacyjno-wyszukiwawczych mylnie jest utożsamiane z polihierarchicznością. Polihierarchia jest wynikiem zastosowania dwóch lub więcej zasad podziału klasy macierzystej (bez względu na metodę jej wyodrębnienia) i odzwiedla na ogół związek hierarchii generycznej między elementami klasy. Klasyfikacją polihierarchiczną może być zarówno klasyfikacja ujawniająca własności kategoryjne przedmiotów, jak i aspekty dziedzinowe lub będąca reprezentacją mieszaną (dziedzinowo-kategoryjną). Należy przy tym pamiętać, że kategorie, zwłaszcza kategorie nieklasyczne, grupując wyrażenia reprezentujące przedmioty określonej rzeczywistości dokumentacyjnej i pozadokumentacyjnej, nie mogą odwoływać się tylko do postulatów znaczeniowych języka naturalnego, ignorując obiektywne i niekiedy subiektywne, własności tej rzeczywistości, choć w praktyce jiw kategorie też bywają porządkowane według relacji hierarchicznych, w tym relacji hierarchii generycznej.

Kategoryzacja słownictwa języków informacyjno-wyszukiwawczych należy do ważnych problemów badawczych. Zajmowali się tym zagadnieniem między innymi: Shiyali Ramamrita Ranganathan<sup>24</sup>, Erie de Grolier<sup>25</sup>, Brian Vickery<sup>26</sup>, Anthony Foskett<sup>27</sup> i Olgierd Ungurian<sup>28</sup>. Kategoryzacja wymaga z jednej strony odwoływania się do wiedzy pozajęzykowej, umożliwiającej wskazanie najistotniejszych dla użytkownika (relewantnych) cech desygnatów, które powinny być odwzorowane w znaczeniu wyrażenia jiw, z drugiej zaś – wskazania cech konotacyjnych i denotacyjnych uznanych za relewantne ze względu na funkcje metainformacyjną i wyszukiwawczą jiw. Dobre systemy kategoryzacji słownictwa jiw zawsze są w jakiś sposób związane ze specyfiką rzeczywistości i języka określonej dziedziny. Podobnie jak w językach naturalnych, również w jiw, kategoryzację jednostek leksykalnych można przeprowadzać biorąc pod uwagę różne ich cechy. W większości języków informacyjno-wyszukiwawczych najogólniejszy podział kategoryjny jednostek leksykalnych jiw opiera się na kryterium syntaktycznym, a dopiero dalsze podziały odwołują się do kryterium semantycznego. Czasami nie jest możliwe rozgraniczenie kategorii semantycznych i syntaktycznych. Ze względu na przedmiot

---

<sup>23</sup> Wereszczyńska-Cisło 1984.

<sup>24</sup> Ranganathan 1967.

<sup>25</sup> Grolier 1962.

<sup>26</sup> Vickery 1997.

<sup>27</sup> Foskett 1996.

<sup>28</sup> Ungurian 1983, 1982, 1976, 1975.

i cele rozprawy dalsze rozważania będą dotyczyć przede wszystkim kategoryzacji semantycznej jednostek leksykalnych jiw jako elementów systemu leksykalnego. W wielu wypadkach konieczne jednak będzie odwoływanie się również do kategorii syntaktycznych, na przykład wtedy, gdy organizacja systemu leksykalnego opiera się na obydwu tych kryteriach jednocześnie. Często bowiem w językach informacyjno-wyszukiwawczych, zarówno uniwersalnych jak i specjalistycznych, przeprowadza się najpierw kategoryzację syntaktyczną leksyki, wyróżniając klasy wyrażen autosyntaktycznych i synsyntaktycznych, a następnie – zależnie od przyjętych kryteriów relewancji – ustala kategorie semantyczne, których wykładnikami w systemie leksykalnym jiw są kategorie leksykalne.

Termin *kategoria* do teorii języków informacyjno-wyszukiwawczych wprowadził Shiyali Ramamrita Ranganathan, wyodrębniając przy tym pięć kategorii, o których będzie dalej mowa. Z analizy użycia tego terminu w tekstach wynika, że w sensie najogólniejszym jest on rozumiany jako klasa obiektów pewnego uniwersum powstała w wyniku jego podziału na podstawie cech uznanych z punktu widzenia pewnych kryteriów relewancji za elementarne. Kategorie, będąc wykładnikami ogólnych znaczeń, do których, na przykład drogą definiowania, dadzą się sprowadzić inne znaczenia wyrażane w danym jiw, wyznaczają główne rodzaje lub typy obiektów kategoryzowanego uniwersum. Olgierd Ungurian<sup>29</sup> za Brianem Vickerym<sup>30</sup> nazywa kategoriami „pojęcia o wysokim stopniu abstrakcji, mające szeroki zakres zastosowania, wybrane przez intelekt w bezpośrednim lub pośrednim nawiązaniu do wiedzy empirycznej i wykorzystywane przez intelekt do celów interpretacji tej wiedzy”. Wynika stąd, że kategorie są to wykładniki (niekoniecznie jawne w planie wyrażania) pewnych, najczęściej szerokich znaczeń, wyróżniających w planie treści jiw obszerne pola semantyczne. Inaczej mówiąc, „kategorie służą do grupowania wyrażen, odwzorowując zwykle podstawowe aspekty rzeczywistości”<sup>31</sup>. Niektóre z tak wyodrębnionych kategorii mają charakter kategorii ontologicznych. Innym zagadnieniem jest kategoryzacja samych wyrażen jiw, którą będziemy się zajmować tylko w takim stopniu i zakresie, w jakim będzie tego wymagać problem kategoryzacji leksykalnej.

W literaturze można spotkać się z opinią, że kategorie powinny mieć charakter uniwersalny, podczas gdy fasety mogą być odbiciem specyfiki opisywanego fragmentu rzeczywistości. Dość radykalnie tezę tę sformułował Ungurian, pisząc, że „aby zapewnić spójność języków informacyjnych, należy używać kategorii uniwersalnych, nie zaś branżowych”<sup>32</sup>. W jakimś sensie do

---

<sup>29</sup> Ungurian 1976, s. 128.

<sup>30</sup> Por. „Concepts of a high degree of generality with a wide area of application elaborated by the mind in referring directly or indirectly to empirical knowledge and utilized by the mind in interpreting such knowledge”. Vickery 1953, s. 54.

<sup>31</sup> Słownik 1993a, s. 69.

<sup>32</sup> Ungurian 1982.



postulatu tego nawiązuje Lucyna Bielicka<sup>33</sup>, proponując zestaw ośmiu kategorii uniwersalnych. U Bielickiej postulat ten ma jednak mniej kategoriyczny charakter. „W każdym z opracowywanych tezaurusów możliwe jest wprowadzenie większej lub mniejszej liczby kategorii (...). Dodatkowe kategorie mogą wystąpić wskutek dzielenia niektórych (...) kategorii na dwie lub więcej. Może się również zdarzyć, że w niektórych tezaurusach wystąpi potrzeba wprowadzenia innych, dodatkowych, nie wymienionych tu kategorii. Mniej kategorii może być wtedy, gdy niektóre z (...) tu wymienionych okażą się w danym tezaurusie nieprzydatne. Niekonieczne jest też zachowanie podanych tu nazw kategorii w formie nie zmienionej. Mogą one być zmodyfikowane i dostosowane do zakresu tematycznego danego tezaury<sup>34</sup>”.

Jako ilustrację przytoczymy obraz organizacji pola semantycznego w *Tezaurusie z zakresu poczty*<sup>35</sup>. Twórcy języka podzielili jego pole semantyczne na trzy części, które można byłoby nazwać podtezaurusami czy tezaurusami składowymi: A. Mechanizacja i automatyzacja poczty; E. Eksploatacja i przewóz poczty; O. Organizacja i ekonomika poczty. W każdej z części wyróżnili następnie maksymalnie pięć kategorii, które nazwali podstawowymi. Były to:

- dziedziny wiedzy i gałęzie gospodarki narodowej<sup>36</sup>
- przedmioty
- osoby
- procesy i czynności
- prawa, wielkości i własności.

„Nazwy powyższych kategorii starano się, w miarę możliwości, formułować w sposób jak najwierniej określający ich zakres tematyczny w poszczególnych działach. Nie wszystkie kategorie wystąpiły we wszystkich działach (...). Tam, gdzie było to potrzebne, ze względu na stosunkowo znaczną liczbę deskryptorów, kategorie podzielono na podkategorie, czyli fasety<sup>37</sup>, uzyskując w efekcie następującą strukturę słownika:

- A. Mechanizacja i automatyzacja poczty
  - A.2. Urządzenia i sprzęt pocztowy
    - A.2.1. Maszyny i urządzenia pocztowe
    - A.2.2. Urządzenia do transportu poczty
    - A.2.3. Meble i sprzęt pocztowy
  - A.4. Czynności
- E. Eksploatacja i przewóz poczty

<sup>33</sup> Bielicka 1988.

<sup>34</sup> Ibidem s. 54.

<sup>35</sup> Tezaurus 1985.

<sup>36</sup> Wyodrębnionego na mocy kryterium dziedzinowego układu klas będących nazwami dziedzin wiedzy i gałęzi gospodarki narodowej nie nazwalibyśmy kategorięzacją a raczej podziałem dziedzinowym lub dyscyplinarnym.

<sup>37</sup> Tezaurus 1985, s. XIV.

- E.2.           Obiekty w eksploatacji pocztowej
- E.2.1.        Druki i dokumenty pocztowe
- E.2.2.        Przesyłki pocztowe
- E.2.3.        Inne obiekty w eksploatacji pocztowej
- E.4.           Czynności w eksploatacji pocztowej
- E.4.1.        Nadawanie, wydawanie i doręczanie przesyłek pocztowych
- E.4.2.        Opracowywanie i przewóz przesyłek pocztowych
- E.5.           Własności i wielkości
- O.            Organizacja i ekonomika poczty
- O.1.          Dyscypliny naukowe i gałęzie gospodarki narodowej
- O.2.          Jednostki organizacyjne
- O.2.1.        Jednostki organizacyjne poczty
- O.2.2.        Systemy organizacyjno-komunikacyjne w poczcie
- O.3.          Użytkownicy i pracownicy poczty
- O.4.          Procesy i czynności
- O.4.1.        Usługi i służby pocztowe
- O.4.2.        Planowanie i kontrola w poczcie
- O.4.3.        Rozwój poczty
- O.5.          Prawa, własności i wielkości
- O.5.1.        Prawa i przepisy pocztowe
- O.5.2.        Wielkości i własności ekonomiczne. Inne cechy

Pojęcie kategorii w językach informacyjno-wyszukiwawczych mocno związane jest z teorią klasyfikacji Ranganathana i samą Klasyfikacją Dwukropkową<sup>38</sup>. W teorii Ranganathana kategorie podstawowe (ang. *fundamental categories*) wykazują funkcjonalne podobieństwo do kategorii ontologicznych, zaś fasety<sup>39</sup> rozumiane są jako „przejawy kategorii w klasie”, wyznaczające w obrębie kategorii podzbiór jednostek leksykalnych należących do określonej dziedziny. Ten sam zbiór słownictwa dzielony jest według dwóch kryteriów: dziedzinowego, prowadzącego do wyodrębnienia klas, i kategoryjnego, prowadzącego do wyodrębnienia kategorii. Takimi własnościami charakteryzuje się między innymi Klasyfikacja Dwukropkowa, której uniwersum zostało podzielone: raz na osi wiedzy zgodnie z zasadami klasyfikacji dziedzin nauki i działalności praktycznej, drugi raz – na osi kategorii. Rezultatem dwukrotnego podziału uniwersum jest tzw. fasetowa organizacja symboli Klasyfikacji Dziesiętnej, polegająca na wyodrębnieniu wśród symboli znakujących klasy i podklasy dziedzin nauki i działalności praktycznej symboli reprezentujących poszczególne kategorie i podkategorie zwane fasetami. Elementami faset są izolaty<sup>40</sup>, czyli symbole reprezentujące znaczenia przyjęte

<sup>38</sup> Pierwotnie używany był tylko jeden wskaźnik – dwukropek – łączący symbole klasyfikacyjne poszczególnych faset w symbolu rozwiniętym i stąd wzięta się nazwa klasyfikacji.

<sup>39</sup> Termin *fasety* został przejęty z terminologii jubilerskiej, gdzie oznacza ścięcie granistej krawędzi naroża szlifowanego kamienia, potęgujące efekty barwne wskutek załamania i rozszczepiania promieni świetlnych. Przenośne użycie tego terminu w teorii i w praktyce podkreśla wieloaspektową analizę rzeczywistości, której elementy podlegają klasyfikacji.

<sup>40</sup> Termin *izolaty* w teorii i praktyce języków informacyjno-wyszukiwawczych używany jest w kilku znaczeniach, z których najważniejsze to: 1) wyrażenie klasyfikacji fasetowej uznane za

w Klasyfikacji Dziesiątej za semantycznie proste. Podział na osi wiedzy jest jawny i tworzy tzw. porządek pionowy, zaś podział na osi kategorii<sup>41</sup> jest ukryty w tym pierwszym, wyznaczając „przejawy kategorii” w poszczególnych klasach dziedzinowych. Wyraża się w postaci wyróżnienia w każdej klasie dziedzinowej, uporządkowanych według relacji generycznej i, rzadziej, me-reologicznej, podzbiorów tej klasy charakteryzujących się wspólnym składnikiem semantycznym, będącym konsekwencją zakwalifikowania do tej samej kategorii. Systematyka leksyki w ramach jednostek należących do tej samej kategorii i tej samej klasy dziedzinowej ma charakter wielostopniowy i polihierarchiczny, co wynika z konsekwentnego stosowania zawsze jednej relacji organizującej. Ranganathan nazwał to wielokrotnym przejawianiem się kategorii w klasie. Wprowadził też rozróżnienie między klasyfikacją fasetową a klasyfikacją analityczno-syntetyczną. W klasyfikacji analityczno-syntetycznej jednostki leksykalne i fasyty są sprowadzane do kategorii ogólnych, uniwersalnych. Jako przykład klasyfikacji analityczno-syntetycznej wskazał, między innymi, stworzoną przez siebie Klasyfikację Dwukropkową. Inny charakter mają klasyfikacje fasetowe, będące schematami empirycznymi, w których nie wyodrębnia się kategorii ogólnych. Zauważmy, że klasyfikacja analityczno-syntetyczna i klasyfikacja fasetowa mogą być traktowane jako dwa gatunki tego samego rodzaju, który można byłoby nazwać *klasyfikacją z kategoriami*. Różnica między nimi polega na tym, że pierwsza z nich operuje kategoriami ogólnymi (uniwersalnymi), druga zaś – kategoriami specjalistycznymi (według terminologii Unguriana – branżowymi). Naszym zdaniem kategorie specjalistyczne, abstrahując od stopnia trudności i pracochłonności tego procesu, powinny dać się wyprowadzić z kategorii ogólnych. Kategorie specjalistyczne można wyróżnić albo poprzez selekcję i uszczegółowienie kategoryzacji uniwersalnej, albo poprzez podejście funkcjonalne, zgodnie z którym kategorie przyjmują postać pól asocjacyjnych ukształtowanych wokół specjalistycznych pojęć centralnych. W praktyce spotyka się też systemy kategorii o charakterze mieszanym. Zatem wszelkie klasyfikacje według kryterium

---

semantycznie proste, a więc najprostsza wyróżniona w tej klasyfikacji klasa; 2) w teorii Ranganathana element uniwersum wiedzy, który rozpatrywany samodzielnie nie jest tematem dokumentu, ale w połączeniu z tematem, reprezentowanym zwykle (w tej teorii) przez nazwę pewnej dziedziny wiedzy lub działalności praktycznej, tworzy tzw. temat rozwinięty, którego zakres jest węższy od zakresu tematu. Ponieważ temat jest najczęściej nazwą pewnej klasy reprezentującej dziedzinę nauki lub działalności praktycznej, izolatami są elementy klasyfikacji reprezentujące pewien obiekt, zjawisko, proces, własność czy relację, które mogą stać się elementami takich klas. Np. „złoto” jest izolatą, która wraz z tematem prostym „chemia” tworzy temat rozwinięty „chemia złota”, „dzieci” to izolata tworząca wraz z tematem prostym „medycyna” temat rozwinięty „pediatria” (medycyna dziecięca), „ciało stałe” to izolata tworząca wraz z tematem prostym „fizyka” temat rozwinięty „fizyka ciała stałego”; 3) również w teorii Ranganathana nazwa rodzajowa dla pewnej treści, jej wykładnika językowego w języku naturalnym i jej wykładnika językowego w języku informacyjno-wyszukiwawczym traktowanych w izolacji, tj. bez odniesienia do struktury języka informacyjno-wyszukiwawczego. Słownik 1993a, s. 55.

<sup>41</sup> Explicite ujawnia się on natomiast na osi syntagmatycznej, konstytuując tzw. formułę fasetową, czyli porządek zapisu wyrażen złożonych.

ustroju pola semantycznego<sup>42</sup> można byłoby podzielić na: klasyfikacje bez kategorii, tj. wyliczające<sup>43</sup> (monodendrytowe, monohierarchiczne, jednonzrębowe) i klasyfikacje z kategoriami (polidendrytowe, polihierarchiczne, wielozrębowe; analityczno-syntetyczne lub fasetowe), pamiętając że w praktyce większość klasyfikacji ma charakter mieszany.

Cechą charakterystyczną Klasyfikacji Dwukropkowej jest zatem jej dwuwymiarowość, przejawiająca się podwójną organizacją paradygmatyki – na osi wiedzy i na osi kategorii. W siódmym<sup>44</sup> wydaniu Klasyfikacji Dwukropkowej, opublikowanym w 15 lat po śmierci Ranganathana, na osi wiedzy wyróżniono następujące klasy główne<sup>45</sup>:

|   |  |
|---|--|
| B | Matematyka (Mathematics)   |
| C | Fizyka (Physics)   |
| D | Inżynieria (Engineering)   |
| E | Chemia (Chemistry)   |
| F | Technika (Technology)  |
| G | Biologia (Biology)   |
| H | Geologia (Geology)   |
| I | Botanika (Botany)  |
| J | Rolnictwo (Agriculture)  |
| K | Zoologia (Zoology)   |
| L | Medycyna (Medicine)  |
| M | Sztuka użytkowa (Useful arts)  |
| Δ | Mistyka i doświadczenia duchowe (Mysticism and spiritual experience) |
| N | Sztuki piękne (Fine Arts)  |
| O | Literatura (Literature)  |
| P | Językoznawstwo (Linguistics)   |
| Q | Religia (Religion)   |
| R | Filozofia (Philosophy)   |
| S | Psychologia (Psychology)   |
| T | Edukacja (Education)   |

---

<sup>42</sup> Termin ten wprowadził do teorii i praktyki języków informacyjno-wyszukiwawczych Ungurian.

<sup>43</sup> Spośród klasyfikacji o przewadze cech klasyfikacji wyliczającej najbardziej znana jest Klasyfikacja Biblioteki Kongresu; charakter wyliczający miała też pierwotnie Klasyfikacja Dziesiętna Deweya.

<sup>44</sup> Pierwsze wydanie tablic Klasyfikacji ukazało się w 1933 roku. W 1937 ukazała się drukiem książka *Prolegomena to library classification* będąca wykładnią poglądów teoretycznych Ranganathana na klasyfikację (Ranganathan 1967). Stała się ona punktem wyjścia do zmodyfikowania pierwszej wersji KD i opublikowania w 1939 r. wydania drugiego. Kolejne ukazały się w latach: 1950, 1952 (pierwsze, w którym Ranganathan zastosował w praktyce teorię kategorii nazwanych fundamentalnymi), 1957, 1960 (poprawiony dodruk w 1963) i 1987 (wydanie siódme). Obecnie KD praktycznie jest wykorzystywana jedynie w kilku bibliotekach w Indiach. Nie umniejsza to jednak roli i wkładu Ranganathana do teorii klasyfikacji i bibliotekarstwa.

<sup>45</sup> Klasy główne cytowane za: Foskett 1996, s. 321-322. Zdaniem Fosketta wydanie siódme Klasyfikacji Dwukropkowej (CC7) ma wiele błędów merytorycznych i drukarskich, czyniących jego użyteczność problematyczną. „CC7 nie jest wyrazem prawdziwej refleksji nad wkładem Ranganathana do bibliotekarstwa” Foskett 1996, s. 321-322.

|   |                                      |
|---|--------------------------------------|
| U | Geografia (Geography)                |
| V | Historia (History)                   |
| W | Nauki polityczne (Political science) |
| X | Ekonomia (Economics)                 |
| Y | Socjologia (Sociology)               |
| Z | Prawo (Law)                          |

Ranganathan zdefiniował klasę główną jako dowolną klasę wyodrębnioną na pierwszym szczeblu podziału uniwersum. Twierdził, że klasy główne powinny mieć charakter klas konwencjonalnych i homogenicznych oraz rozłącznych i wyczerpujących, z czego wynika, że starał się stosować w praktyce dwie nadrzędne zasady – zgodności z przyzwyczajeniami użytkowników oraz podziału logicznego. W opublikowanych pracach dawał też wyraz przekonaniu o niestabilności jakiegokolwiek sekwencji klas głównych. Dlatego wysiłki koncentrował nie na formułowaniu uzasadnienia teoretycznego dla porządku klas głównych wyodrębnionych w Klasyfikacji Dwukropkowej, a na definiowaniu zasad kategoryjnej organizacji pojęć wewnątrz klas głównych oraz na analizie interdyscyplinarnych związków pojęciowych.

Na osi kategorii Ranganathan wyróżnił w Klasyfikacji Dwukropkowej pięć kategorii podstawowych:

- [P] Indywidualność (ang. Personality)
- [M] Materia (ang. Matter)
- [E] Energia (ang. Energy)
- [S] Przestrzeń (ang. Space)
- [T] Czas (ang. Time).

Kategorie Ranganathana przypominają w pewnym stopniu kategorie ontologiczne wyróżniane w filozofii europejskiej, jednak wykazanie jednoznaczności „pokrewieństwa” tych dwóch systemów kategoryjnych byłoby dość trudne, jeśli w ogóle możliwe<sup>46</sup>. Być może przyczyny należy szukać w odmienności, hinduskiej filozofii poznania i filozofii w ogóle lub w tym, że kategorie w Klasyfikacji Dziesiątej, o czym pisał sam Ranganathan, zostały wyłonione w efekcie analizy treści dokumentów i tworzących ją pojęć, nie zaś ontologicznych własności elementów uniwersum. Ranganathan zdefiniował *kategorię podstawową* jako nazwę rodzajową oznaczającą kategorie Czas, Przestrzeń, Energia, Materia i Indywidualność. Z kolei znaczenia poszczególnych kategorii zostały określone przez wskazanie zaliczonych do nich typów obiektów. I tak:

– do kategorii Indywidualności [P] należą jednostki leksykalne określające osobę ludzką oraz wszystko, co ma właściwości indywidualne, np. zwierzę, roślinę, ciało nieorganiczne, języki, religie itp., byty konkretne i społeczne, byty abstrakcyjne itp. Kategoria Indywidualności jest na przykład jedyną kate-

<sup>46</sup> Foskett 1996; Iyer 1995; Ungurian 1976, 1975.

gorią wyodrębnianą w klasie O Literatura, ale w czterech różnych „przejawach”: [P1] (język), [P2] (forma), [P3] (autor) i [P4] (dzieło);

– do kategorii Materia [M] należą jednostki leksykalne określające między innymi: materiały naturalne, materiały prawie naturalne, katalizatory, półfabrykaty używane jako surowce, produkty i wyroby końcowe, usługi bezpośrednie i pośrednie, właściwości, stany, atrybuty bytów, w tym wartość, czyli atrybut z punktu widzenia stopnia doskonałości, zasad, jakości lub czegoś, co jest istotnie pożądane, np. struktura cylindryczna, kolor czerwony, właściwości utleniające, wartość etyczna, pedagogiczna, społeczna itp.;

– do kategorii Energia [E] należą jednostki leksykalne określające wszelkie działanie, które może być wykonane i któremu może podlegać byt, między innymi: fizyczne działanie człowieka na człowieka lub na materiał (badanie, poznawanie, przetwarzanie, negocjowanie itp.), oddziaływanie intelektualne, wzajemne oddziaływanie między obiektami fizycznymi lub biologicznymi, wzajemne oddziaływanie pomiędzy układami naturalnymi w przyrodzie lub grupami społecznymi, autoakcje wewnątrz obiektu, metody działania wykorzystywane we wzajemnym oddziaływaniu lub autoakcji itp.;

– do kategorii Przestrzeń [S] należą jednostki leksykalne określające przedmioty związane z przestrzenią, między innymi: powierzchnią Ziemi, przestrzenią wewnątrz Ziemi, atmosferą ziemską i wszechświatem, a także jednostki określające wszelkie względne umiejscowienia przedmiotów itp.;

– do kategorii Czas [T] należą jednostki leksykalne określające przedmioty związane z czasem, między innymi: daty, okresy trwania, częstotliwość, epoki, okresy związane z ruchem, okresy meteorologiczne (pora deszczowa, sucha itp.).

Konsekwencją zastosowanej przez Ranganathana metody tworzenia kategorii jest ich specyficzny charakter. Jest to dobrze widoczne na przykładzie kategorii Indywiduum odnoszącej się do tego, co najistotniejsze w danym obszarze rozważań, co stanowi ich obiekt centralny. Elementy tej kategorii identyfikowane są na podstawie roli, jaką pełnią w badanej dyscyplinie, na przykład **Stal**, która w dziale Przemysł budowlany zaliczona jest do kategorii Materia, w dziale Hutnictwo należy do kategorii Indywiduum. Praktyka wykazała, że zaproponowany przez Ranganathana podział na kategorie nie jest precyzyjny, zaś ich liczba – niewystarczająca. Ranganathan sam to przyznał, wielokrotnie dokonując pewnych modyfikacji. W ostatnim (siódmym) wydaniu Klasyfikacji Dwukropkowej dokonano na przykład podziału kategorii Materia na podkategorie: Materia [M], Materia – własność [MP] i Materia – metoda [MM], przenosząc jednocześnie pewne izolaty (na przykład **Gramatyka** z klasy Językoznawstwo), które wcześniej były elementami kategorii [E] do kategorii [MP]. Można zatem powiedzieć, że kategorie w Klasyfikacji Dwukropkowej mają charakter pewnej ogólnej struktury, mogącej podlegać zmianom, rozwinięciom, uściśleniom itd. w rezultacie zmian w uniwersum i wniosków płynących z praktyki klasyfikowania i wyszukiwania dokumentów.

Problematyką kategorii i kategoryzacją intensywnie zajmowała się też Classification Research Group (CRG). W pracach CRG przez kategorię rozumie się pewną ogólną klasę znaczeniową, często wyodrębnioną na podstawie cech dziedzinowych (dyscypliny). Faseta jest z kolei zbiorem jednostek leksykalnych (zorganizowanym najczęściej według relacji generycznej) obejmującym nazwy przedmiotów, procesów, zjawisk, stanów, własności i innych zagadnień szczegółowych. Różnice między rozumieniem kategorii i fasety przez Ranganthana i przez CRG wynikają zapewne z faktu, że języki informacyjno-wyszukiwawcze, będące przedmiotem zainteresowania CRG, były językami specjalistycznymi (o ograniczonym zakresie), na przykład: bibliotekoznawstwo, oświata, sport i wychowanie fizyczne, podczas gdy klasyfikacja Ranganthana miała charakter uniwersalny. Wyodrębnianie w językach nieuniwersalnych kategorii odwzorowujących pewien podstawowy podział rzeczywistości uznano za nieuzasadnione ze względów praktycznych. W klasyfikacjach specjalistycznych na ogół też nie przeprowadza się niezależnego od kategoryalnego podziału słownictwa według kryterium dziedzinowego, co sprawia, że tworzone w nich fasety przyjmują postać albo drzew klasyfikacyjnych przedmiotów, procesów, zjawisk itd. charakterystycznych dla danej dziedziny, albo drzew klasyfikacyjnych dyscyplin i zagadnień badawczych zawierających się w danej dziedzinie, albo też hybrydową postać drzew klasyfikacyjnych dziedzinowo-przedmiotowych, w których klasy pierwszego czy drugiego poziomu podziału zajmują jednostki leksykalne odnoszące się do pewnych dziedzin lub zagadnień badawczych w nich wyróżnionych, natomiast dalszy podział przebiega według relacji hierarchii tematycznej, pozwalającej wyodrębnić przedmioty badań, metody i techniki badawcze, podmioty badań itd., porządkowane w dalszej kolejności według relacji generycznej lub mereologicznej.

Cechą wspólną specjalistycznych podziałów kategoryalnych jest ich bliski związek z rzeczywistością. Przykładami specjalistycznych modeli kategoryalnych są między innymi systemy Aliny i Briana Vickerych oraz Olgierda Unguriana. Obydwa stanowią rezultat szczegółowej analizy podstawowych kategorii ontologicznych, ich odwzorowania w semantyczne kategorie słownictwa nauki i techniki oraz wyodrębnienia kategorii specjalistycznych specyficznych dla określonej dziedziny.

Alina i Brian Vickery, projektując specjalistyczny język informacyjno-wyszukiwawczy<sup>47</sup>, wyróżnili następujące kategorie:

- I Przedmioty, substancje, obiekty materialne
  - Powstałe naturalnie, np. minerał, zwierzę, gleba, roślina
  - Produkty
  - Narzędzia
  - Konstrukty mentalne (umysłowe), np. figury geometryczne, równania, wzory
- II Ich części

---

<sup>47</sup> Vickery 1987a.

- Elementy składowe
- Organy
- III Układy przedmiotów
- IV Atrybuty przedmiotów
  - Jakości, właściwości
  - Struktury
  - Miary
  - Procesy, zachowania
- V Obiekty działania (pacjensy)
- VI Związki między przedmiotami, interakcje
  - Rezultaty
- VII Operacje na przedmiotach, działania na rzeczach
  - Eksperymentalne
  - Mentalne
- VIII Wykorzystanie rzeczy
- IX Miejsce, warunki, otoczenie
- X Czas

Problematykę tę kontynuowała Alina Vickery, tworząc model języka dla specyficznie rozumianego, bo niebezpośredniego (skierowującego), eksperymentalnego systemu ekspertowego z dziedziny ogrodnictwa Plexus<sup>48</sup>. Baza systemu Plexus liczy zaledwie 500 charakterystyk publikacji, instytucji i ekspertów z dziedziny ogrodnictwa, zaś indeks wyszukiwawczy – 2500 kluczy. Pytania użytkowników sformułowane w języku naturalnym są przez system, po ewentualnym zapytaniu o dodatkowe informacje, przekształcane do postaci instrukcji, a następnie użytkownikowi wskazywane są potencjalne źródła odpowiedzi. W systemie Plexus wykorzystano mechanizm kategoryzacji gramatycznej oparty na teorii przypadków Fillmore'a. Wyróżniono następujące kategorie leksykalne:

|                    |   |
|--------------------|---|
| Obiekt             | (roślina, mikroorganizm, choroba itp. – np.: róża, mszyce)        |
| Część obiektu      | (np. łodyga, korzeń)  |
| Operacje           | (działania wykonywane przez ludzi na obiektach – np. plewienie)   |
| Proces             | (działania obiektów – np. kwitnienie)                             |
| Interakcja         | (między obiektami – np. A atakuje B: mszyce atakują róże)         |
| Instrument         | (stosowany w działaniach wykonywanych na obiektach – np. sekator) |
| Atrybut            | (określający rośliny, glebę, działania itp. – np. stałozielone)   |
| Środowisko obiektu | (np. zbiornik wodny)  |

<sup>48</sup> Vickery 1987b, c.



Wykorzystanie obiektów (np. dekoracyjne)  
Czas (ogólne i specyficzne określenia czasu,  
jak np. pory roku itp.)

Lokalizacja geograficzna

W kategoriach tych wyróżniono podkategorie, na przykład kategoria Obiekt została podzielona na:

- Obiekt\_1 – Rośliny
- Obiekt\_2 – Zwierzęta
- Obiekt\_3 – Mikroorganizmy
- Obiekt\_4 – Choroby lub symptomy
- Obiekt\_5 – Gleby lub skały
- Obiekt\_6 – Nasiona

Każde wyrażenie języka systemu Plexus powiązane jest dodatkowo z odpowiednimi kategoriami i klasami języka BSO (Broad System of Ordering), wyznaczającymi podział dyscyplinarny dziedziny. Ramy semantyczne są tu oparte na kombinacji subkategorii powiązanych z typem kategorii. Pytania użytkowników „dopasowuje się” do poszczególnych ram w sposób analogiczny do tego, jak to ma miejsce w przypadku formuł fasetowych. Na przykład, kiedy użytkownik prosi o informacje na temat „Róż”, najpierw na podstawie stop-listy system eliminuje z pytania wyrażenia, które nie mają wartości wyszukiwawczej, a następnie każde z pozostałych sprawdza w słowniku rdzeni tematycznych. W charakteryzowanym przypadku system odnajdzie artykuł słownikowy rdzenia tematycznego terminu Róże. Informacje w nim zawarte wskazują, że termin ten należy do podkategorii Obiekt\_1. Na tej podstawie „ściągnie” wzorcową ramę podkategorii Obiekt\_1 i wpisze do niej uzyskane dotąd informacje: numer ramy Róże. Kolejnym krokiem jest przepisanie do ramy informacji zawartych w artykule słownikowym terminu, tj. symbolu BSO, nazwy kategorii semantycznej, listy synonimów. Dodatkowo system dopisze numer pozycji terminu w pytaniu. Na podstawie tak zebranych informacji Plexus buduje instrukcję i wybiera optymalną strategię wyszukiwawczą.

Uniwersalne modele kategoriałne szczególną rolę pełnią w językach informacyjno-wyszukiwawczych typu kodu semantycznego, wykorzystujących zasadę definiowania pojęć na podstawie analizy składnikowej znaczeń. Jednostki leksykalne kodów mają strukturę semantycznie umotywowaną, będąc połączeniami wyrażen reprezentujących pojęcia proste wyróżnione w strukturze semicznej języka. Słownik kodu semantycznego rejestruje kody reprezentujące pojęcia uznane za elementarne. Pojęcia te są grupowane w ogólniej zdefiniowanych klasach, które – mimo że języki te najczęściej były budowane dla specjalistycznych systemów informacyjno-wyszukiwawczych – mają zwykle charakter uniwersalnego, ontologicznego systemu kategoriałnego. Przykładem bardziej znanego kodu semantycznego jest Język ASM-WRU zwany też Kodem Semantycznym Perrye’go i Kenta.

Język ASM-WRU<sup>49</sup>, zbudowany w latach 1954-1959 w Western Reserve University (WRU) Center for Documentation and Communication Research (USA) przy współdziałaniu Madeline Barry, przeznaczony był do stosowania w stworzonym w WRU, na zlecenie American Society for Metals (ASM), systemie informacji metalurgicznej. W odróżnieniu od takich jiw jak na przykład UKD czy język haseł przedmiotowych, kody semantyczne często były budowane do realizacji funkcji informacyjnej, właściwej językom naturalnym, a ponadto funkcji metainformacyjnej i wyszukiwawczej, analogicznie do tzw. metajęzyków semantycznych stosowanych w lingwistyce do opisu znaczeń wyrażen języka naturalnego<sup>50</sup>. Kod Perry'ego i Kenta ma pole semantyczne ograniczone do metalurgii i założoną funkcję metainformacyjną (i wyszukiwawczą). Zasady strukturalizacji wiedzy, przyjęte w tym języku, mają charakter uniwersalny. Wpływ specjalistycznego charakteru systemu informacyjno-wyszukiwawczego, na użytek którego kod był stworzony, widać w doborze i nazewnictwie podkategorii niższych poziomów.

W słowniku kodu ASM-WRU ogólne cechy semantyczne obiektów, uznane za elementarne (prymitywy semantyczne), reprezentowane przez 214 tzw. mnożników semantycznych, podzielono na sześć, ulegających dalszemu podziałowi, kategorii semantycznych:

- I Pojęcia ogólne:
  - Idee, pojęcia, teorie
  - Relacje ogólne
  - Własności ogólne
  - Obszary badań lub działalności
  - Siły
  - Klasyfikacje
- II Stosunki:
  - Ekonomiczne
  - Fizyczne
  - Społeczne
    - podstawowe pojęcia socjologiczne
    - pojęcia dotyczące społeczeństwa
    - procesy zachodzące w społeczeństwie
    - wskaźniki dotyczące społeczeństwa
- III Stany:
  - Psychologiczne
  - Wyzwalające działanie
  - Fizyczne
    - odczuwalne zmysłowo

---

<sup>49</sup> Innymi znanymi kodami semantycznymi był ukraiński język *RX*-kodów Skorochoďki. W Polsce prace nad budową kodu semantycznego prowadzili w latach siedemdziesiątych Bożenna Bojar i Olgierd Wojtasiewicz.

<sup>50</sup> Bojar 1979.

- kondycjonalne
- własności wywołujące stany

#### IV Procesy:

- Fizyczne
- Ogólne
- Materialne
  - powodujące zmiany
  - konstruktywne
  - destruktywne
- Społeczne i przemysłowe

#### V Substancje:

- Ogólne
- Szczególne (organiczne, nieorganiczne)
  - organiczne w przyrodzie
  - nieorganiczne wytworzone przez człowieka
- Preparaty, przetwory

#### VI Przedmioty:

- Ogólne (naturalne, artefakty)
  - naturalne
  - produkty, artefakty
- Specjalne
  - naturalne
  - produkty, artefakty

Elementarnymi jednostkami leksykalnymi kodu o funkcji autosyntaktycznej, będącymi zarazem językowymi nośnikami znaczeń szczegółowych, są tzw. deskryptory proste, składające się z mnożników semantycznych i wykładników jednej z dziesięciu relacji paradygmatacznych (zwanymi relacjami analitycznymi). Deskryptory proste są więc ciągami czteroznakowymi. Trzy z tych znaków są wykładnikami przynależności deskryptora do określonej kategorii semantycznej, np. M\_TL – ‘bycie metalem’, czwarty znak, w postaci infiksu, służy do wyrażania relacji między kategorią a oznaczanym obiektem, np. MATL – ‘być typowym reprezentantem metali; należeć do metali’. Wykładnikami relacji analitycznych są litery A, E, I, O, U, Q, W, X, Y, Z. Ich funkcją jest precyzowanie znaczenia poprzez wskazanie rodzaju związku między znaczeniem reprezentowanym przez mnożnik semantyczny i znaczeniem reprezentowanym przez deskryptor prosty, inaczej mówiąc, związku między oznaczanym przedmiotem a kategorią, do której został zaliczony. Wykładniki te są wyodrębnione jako niezależne kategorie leksykalne, nie podlegające dalszemu podziałowi, wyznaczające strukturę paradygmataczną języka.

- A- – być typowym reprezentantem,
- E- – być składnikiem, produktem materiału,

- I- – być częścią,
- O- – być złożonym z kilku elementów danej klasy,
- U- – być producentem,
- Q- – być użytkownikiem czegoś, być zdeterminowanym przez coś,
- W- – być środkiem, przedmiotem działania,
- X- – nie posiadać jakiejś cechy,
- Y- – być istotną cechą obiektu,
- Z- – posiadać pewne cechy nie reprezentowane przez mnożnik.

Pozostałe relewantne informacje wyrażane są przez sufiksy cyfrowe dołączane do deskryptorów prostych, będące równocześnie numerami wyrażen związanych relacją generyczną z reprezentującym przedmiot deskryptorem, na przykład FZSH.17 – „delfin”. Sufiksy te są więc wykładnikami zakresów znaczeniowych wyrażen kodu. Inną metodą precyzowania cech dystynktywnych przedmiotu jest konkatencja deskryptorów prostych.

W kodzie ASM-WRU przeprowadzono także kategoryzację wyrażen reprezentujących relacje syntagmatyczne (syntetyczne, w późniejszych opracowaniach zwane wskaźnikami roli) niezbędne w opisie sytuacji pozajęzykowych, w których występują desygnaty deskryptorów prostych, lub komunikowaniu aspektu ujęcia tych desygnatów w dokumencie. Pełniąc funkcję podobną do określników w języku haseł przedmiotowych, są wykładnikami relewancji interpretacyjnej. Wyróżniono cztery klasy wykładników relacji syntetycznych, zależnie od łączliwości syntaktycznej z deskryptorami należącymi do określonych kategorii ogólnych, a więc łączące się z oznaczeniami: **materiałów, procesów, własności, warunków**.

Klasy te nie są rozłączne, występują bowiem wykładniki takich relacji syntetycznych, które mogą łączyć zarówno materiały, jak własności, procesy czy warunki. W planie wyrażania są one reprezentowane przez trzyliterowe symbole, zaczynające się literą K i kończące „,” (kropką), na przykład: KOV., KAP., KAI. Początkowo było ich 24, następnie ograniczono się do 14. Kategoryzacja wykładników relacji syntetycznych wykazuje pewne analogie do kategoryzacji mnożników semantycznych.

Cechą charakterystyczną kodu Perry’ego i Kenta jest notacja mnemoniczna korzystająca z mnemoniki rdzeniowej, upodabniającej wyrażenia kodu do wyrażen języka angielskiego, na przykład: M-SR (measure), T-MP (temperature), M-CH (machine), N-ML (animal).

Analizując pojęcie kategorii, nie można pominąć problemu wielowymiarowości pola semantycznego jiw. Liczba proponowanych wymiarów wynosi od dwóch (przedmiotu i treści u Ranganthana) do pięciu (cech formalnych, treści, przedmiotu, czasu i przestrzeni) u Unguriana. W pracach dotyczących konstrukcji języków informacyjno-wyszukiwawczych zwraca się uwagę na przydatność rozróżniania wyrażen tylko pozornie desygnujących te same obiekty. Na przykład, w językach haseł przedmiotowych istnieją tematy nazywające

rodzaje i typy dokumentów, np.: kazania, słowniki, antologie, jako przedmioty treści obok równokształtnych tematów (zwanymi formalnymi) nazywających formy katalogowanych dokumentów.

Koncepcję wielowymiarowości jiw wysunął Ranganathan, wyróżniając w Klasyfikacji Dwukropkowej wymiar przedmiotu i wymiar treści. W piśmiennictwie polskim rozwinął ją Ungurian<sup>51</sup>, opracowując koncepcję języka informacyjno-wyszukiwawczego z zakresu nauk ścisłych i technicznych. Koncepcja Unguriana może być traktowana jako rozwinięcie systemu kategorii Ranganathana, poprzez wyróżnienie i rozwinięcie zestawu wymiarów opisu dokumentów. Ranganathan zaprojektował tylko dwa wymiary – wymiar treści, ujęcia czy aspektu prezentacji określonych fragmentów rzeczywistości w tekście dokumentu, znajdujący wyraz w postaci podziału dyscyplinarnego i wymiar przedmiotu treści, zorganizowany za pomocą podziału kategorialnego. Pojęcia dotyczące cech formalnych dokumentu, takich jak na przykład forma wydawnicza czy język, zgrupowane są u Ranganathana w odrębnych zestawieniach, nazwanych izolatami wspólnymi, funkcjonujących podobnie jak podziały wspólne np. w Uniwersalnej Klasyfikacji Dziesiętej. Ungurian nadał im status odrębnych wymiarów. Konsekwencją zasady monopozycyjności jest u Unguriana wyróżnienie odrębnych wymiarów Czasu i Przestrzeni poza wymiarem Treści. Czas i Przestrzeń mogą być zatem wykorzystane zarówno w celu wskazania lokalizacji w czasie i przestrzeni elementów treści dokumentu, jak i do identyfikacji cech pozat treściowych dokumentu. Zarówno pięć wyodrębnionych przez Unguriana wymiarów jiw, jak i organizacja zastosowana wewnątrz wymiarów (z wyłączeniem wymiaru cech formalnych), mają charakter uniwersalny. Przejawem kategoryzacji specjalistycznej jest rozbudowa wymiaru Przedmiot dokumentu. Opiera się ona na uniwersalnym, dychotomicznym podziale pojęć na nominalne (reprezentujące obiekty jako pewne całości) i predykatywne (reprezentujące relacje), a następnie, również uniwersalnym i dychotomicznym, podziale każdej z klas na: pojęcia nominalne – reprezentujące przedmioty indywidualne, pojęcia nominalne – reprezentujące wieloelementowe zbiory przedmiotów, pojęcia predykatywne – reprezentujące relacje jednoargumentowe (cechy) i pojęcia predykatywne – reprezentujące relacje wieloargumentowe (zmiany). Dalsze kryteria wyznaczania kategorii szczegółowych zdefiniowane zostały również ogólnie, co może sugerować, że ostatecznie uzyskana kategoryzacja powinna mieć walor uniwersalności. Ungurian przetestował swoją koncepcję na dokumentach z zakresu nauk ścisłych i technicznych, stąd traktowana jest jako specjalistyczna, mogąca zdaniem samego autora, okazać się niewystarczająca czy nieadekwatna dla potrzeb systemów o innym zakresie.

Ungurian wyróżnił pięć następujących wymiarów, wyodrębniając w każdym z nich kategorie specyficzne dla danego wymiaru:

---

<sup>51</sup> Ungurian 1983, 1982.

- I. Wymiar cech formalnych, w którym gromadzi się jednostki leksykalne wskazujące własności prezentacji treści dokumentów, podzielone na kategorie:
  1. Forma wydawnicza
  2. Gatunek dokumentu
  3. Przeznaczenie
  4. Język dokumentu;
  
- II. Wymiar treści dokumentu, w którym występują jednostki leksykalne wskazujące aspekty przedmiotów rozważanych w dokumencie, podzielone na trzy kategorie:
  5. Dziedzina wiedzy
  6. Gałąź gospodarki
  7. Sfera działalności ludzkiej;
  
- III. Wymiar przedmiotu obejmujący jednostki leksykalne wskazujące przedmioty treści dokumentów, podzielone na Relacje i Obiekty, które dzielą się dalej na:
  - Relacje
    - Zmiany
      8. Zjawiska
      9. Procesy
    - Cechy (atrybuty zmiany)
      - Stany (obiektów)
        10. Atrybuty zmian
        11. Atrybuty istniejące
        12. Atrybuty potencjalne
  - Obiekty
    - Obiekty zbiorowe
      13. Układy (spontaniczne)
      14. Organizacje (celowe)
    - Obiekty elementarne
      - Naturalne
        15. Instrumenty (naturfakty)
        16. Materiały (naturfakty)
      - Sztuczne
        17. Instrumenty (artefakty)
        18. Materiały (artefakty)
  
- IV. Wymiar czasu, w którym wyróżnionych jest pięć kategorii:
  19. Czas chronologiczny
  20. Czas względny
  21. Czas fenomenologiczny
  22. Podział czasu
  23. Częstotliwość;

- V. Wymiar przestrzeni, w którym wyróżnionych jest sześć kategorii jednostek leksykalnych:
24. Podział geograficzny polityczny
  25. Podział geograficzny fizyczny
  26. Rejonizacja
  27. Formy geograficzne
  28. Orientacje
  29. Lokalizacja względna

Zdaniem Unguriana uzasadnieniem wyróżnienia samodzielnych wymiarów Czasu i Przestrzeni jest fakt, że zgromadzone w nich jednostki leksykalne mogą nazywać różne własności dokumentów, choć najczęściej używane są do oznaczania lokalizacji czasowej i przestrzennej treści dokumentu. Czas może być przedmiotem treści dokumentu traktującego o jakimś nowym podziale czasu, może też stanowić uściślenie opisywanego zdarzenia lub wskazywać cechę formalną dokumentu; podobnie kategorie przestrzenne mogą stanowić przedmiot dokumentu, przestrzenny aspekt omawianego przedmiotu lub lokalizację geograficzną wydania dokumentu. Stąd traktowane są jako wymiary specyficznych charakterystyk dokumentów. Mogą też, w postaci tzw. faset ruchomych, pojawiać się w innych wymiarach. Możliwość przemieszczania się pewnych kategorii z jednych wymiarów do innych występuje także w przypadku kategorii wymiaru Treści, które mogą się pojawiać w wymiarze Przedmiotu. Tak rozumiana „ruchomość” cechuje również inne kategorie, np. kategorię formy wydawniczej czy kategorię gatunku dokumentu.

Ungurian nie traktował zaproponowanego „szkieletu organizacyjnego” jiw jako schematu zamkniętego, zaznaczając, że zarówno liczba wymiarów, jak i wyróżnionych w nich kategorii może być zmieniana. „Wydaje się, że liczba tych wymiarów jest i zawsze pozostanie sprawą dyskusyjną, gdyż trudno czysto teoretycznie udowodnić potrzebę takiej, a nie innej liczby wymiarów, wskazując, jakie to są wymiary. Zawsze liczba tych wymiarów będzie zależeć od tego, jak widzi projektowany język informacyjny jego twórca, i w jakim stopniu poznał on rzeczywiste potrzeby swoich przyszłych klientów, co stanowi najwyższy i najistotniejszy sprawdzian każdego języka informacyjnego”<sup>52</sup>.

Koncepcja wielowymiarowości zawarta jest implicite w konstrukcji wielu języków informacyjno-wyszukiwawczych. W praktyce jednak nie jest stosowana konsekwentnie. Wyróżniane są zwykle dwa lub trzy wymiary (najczęściej przedmiotu, treści i formy), przy czym część leksyki większości jiw może występować w roli wykładników różnych własności dokumentów mimo częściowego podziału ich pola semantycznego na wymiary.

---

<sup>52</sup> Ungurian 1982, s. 37.

Wykorzystując wszystkie przedstawione do tej pory podstawy metodologiczne kategoryzacji, w tym kategoryzacji w językach informacyjno-wyszukiwawczych, przedstawimy charakterystykę wybranych jiw. Będą to: Uniwersalna Klasyfikacja Dziesiętna, język haseł przedmiotowych w ujęciu Adama Łysakowskiego i jhp KABA oraz język deskryptorowy, którego słownikiem jest Tezaurus technologii żywności i przemysłu spożywczego. Trzy pierwsze języki są powszechnie stosowane w bibliotekach i innych instytucjach informacji w Polsce; czwarty zaś zwraca uwagę wzorową i przemyślaną konstrukcją. Podejmiemy też próbę scharakteryzowania kategorii informacyjno-wyszukiwawczych, z jakimi ma do czynienia użytkownik wyszukiwarki internetowej AltaVista. Zrezygnowaliśmy z większej liczby przykładów, bo celem pracy jest identyfikacja pewnych ogólnych stwierdzeń i postulatów, nie zaś wyczerpujący opis systemów kategorii w istniejących językach informacyjno-wyszukiwawczych. Ten element badań kategoryzacji ma stosunkowo obszerną literaturę<sup>53</sup>. Nasza uwaga będzie się koncentrować głównie na kategoriach leksykalnych. Wiele interesującego i wartościowego materiału na temat kategorii gramatycznych i syntaktycznych w jiw, choć przedstawionych w innym ujęciu, można znaleźć w niepublikowanej pracy doktorskiej Ewy Chmielewskiej-Gorczyca<sup>54</sup>.

Charakterystyką obejmujemy obiekty rzeczywistości istniejącej poza językiem informacyjno-wyszukiwawczym (pamiętając o ich dwoistym charakterze – dokumenty i obiekty przedstawione w dokumentach) oraz w jakimś stopniu obiekty samego jiw, zakładając, że jeśli istnieje paralela pomiędzy kategoryzacją obiektów języka naturalnego i dokonującą się za pośrednictwem tego języka kategoryzacją rzeczywistości pozajęzykowej, to powinien też istnieć analogiczny związek między obiektami jiw i obiektami rzeczywistości poza jiw. Z naszych analiz i obserwacji wynika, że nadal jednym z najłabszych elementów teorii i praktyki języków informacyjno-wyszukiwawczych jest niewystarczająca wiedza o bezpośredniej rzeczywistości pozajęzykowej jiw, czyli o rzeczywistości (świecie) dokumentów. Wiedza o bezpośredniej rzeczywistości pozajęzykowej jiw powinna być, obok wiedzy o rzeczywistości opisywanej w dokumentach, jedną z dwóch podstaw orzekania o strukturze i języku informacyjno-wyszukiwawczym w ogóle. Obecnie zauważa się wyraźną dominację drugiej z nich, i to często w postaci wątpliwych, subiektywnie ukształtowanych obrazów. Zapomina się, że kategoryzacja znaczeniowa słownictwa wymaga zawsze z jednej strony odwołania się do wiedzy pozajęzykowej, umożliwiającej wskazanie najistotniejszych dla użytkownika cech desygnatów, które powinny być odwzorowane w strukturze morfologicznej wyrażenia lub w jego definicji, z drugiej zaś – wskazania cech uznanych za relewantne ze względu na funkcje metainformacyjną i wyszukiwawczą jiw – oraz że nie da się poprawnie i wyczerpująco wskazać owych cech relewantnych

---

<sup>53</sup> Na przykład: Artowicz 1997; Grolier 1962; Sosińska 1995, 1993, 1988; Ścibor 1982; Vickery 1975, 1953.

<sup>54</sup> Chmielewska-Gorczyca 1981.



bez odwoływania się do wiedzy o rzeczywistości funkcjonowania jiw. Znaczenie bowiem to nie wierne odzwierciedlenie cech odpowiadającego wyrażeniu obiektu, ale świadoma transformacja językowa uwydatniająca jedne cechy, a pomniejszająca czy wręcz eliminująca inne.

#### KATEGORIE W UNIWERSALNEJ KLASYFIKACJI DZIESIĘTNEJ

W Uniwersalnej Klasyfikacji Dziesiętnej (UKD) kategoryzacja leksykalna wiąże się z kategoryzacją gramatyczno-syntaktyczną jednostek leksykalnych. Stanowi rozwinięcie podziału leksyki UKD na symbole główne i pomocnicze, wśród których lokują się symbole poddziałów wspólnych, symbole poddziałów analitycznych i symbole syntetyczne.

Symbole główne, przeznaczone do wyrażania treści dokumentów, stanowią podstawowy zrąb leksyki UKD, będący schematem klasyfikacyjnym dziesięciu działów dziedzinowych, prezentowanym w słowniku UKD zwanym tablicami głównymi. Każdy symbol główny przynależy do jednej z klas tworzących dziesięć głównych działów UKD. W poszczególnych klasach symbole wyrażające dyscypliny szczegółowe, zagadnienia, metody i przedmioty badawcze, ich elementy itp. uporządkowane są hierarchicznie według relacji generycznej, relacji mereologicznej lub relacji hierarchii tematycznej. Podział dyscyplin w UKD został przejęty przez twórców UKD, Paula Otleta i Henriego La Fontaine'a, z Klasyfikacji Dziesiętnej Deweya, która z kolei pozostawała pod wpływem opracowanej dla zbiorów Public School Library w Saint Louis Klasyfikacji Książek Harrisa. Organizacja tablic głównych UKD, opisywana w literaturze między innymi przez Ranganathana, A. C. Fosketa, Unguriana, Ścibora i Sosińską, nie będzie przedmiotem naszego zainteresowania, ponieważ jest to zagadnienie przede wszystkim z dziedziny klasyfikacji nauki, wiedzy i piśmiennictwa, nie zaś kategoryzacji. Chcielibyśmy jedynie zwrócić uwagę, że zasadniczy zrąb Uniwersalnej Klasyfikacji Dziesiętnej został utworzony ponad 100 lat temu. Niektóre jego fragmenty opisują strukturę nauki znacznie odbiegającą od jej kształtu współczesnego. „Prowadzone przez FID, a obecnie przez Konsorcjum ds. UKD, prace nad aktualizacją klasyfikacji mają na celu dostosowanie jej struktury do wymagań rozwoju nauki, jednakże skomplikowana procedura rewizji schematu UKD, coraz większe trudności znalezienia specjalistów gotowych podjąć się rewizji poszczególnych działów i równocześnie ogromna dynamika zmian zachodzących we współczesnej nauce powodują nienadążanie aktualizacji wielu części UKD za ewolucją reprezentowanych przez nie dyscyplin”<sup>55</sup>. Opis przebiegu prac nad rewizją UKD na przykładzie działu 8 poświęconego językoznawstwu, językom i literaturze pięknej, trwających od 1974 do 1992 r., można znaleźć w artykule Ścibora<sup>56</sup>.

<sup>55</sup> Sosińska 1993, s. 92.

<sup>56</sup> Ścibor 1995.

Obecnie w UKD obowiązuje następujący podział wiedzy na pierwszym poziomie klasyfikacji:

- Dział 0 Zagadnienia ogólne
- 1 Filozofia. Psychologia
  - 2 Religia. Teologia. Religioznawstwo
  - 3 Nauki społeczne. Prawo. Administracja
  - 4 \_\_\_\_\_ (nie wykorzystany)
  - 5 Matematyka. Nauki przyrodnicze
  - 6 Nauki stosowane. Medycyna. Nauki techniczne. Rolnictwo
  - 7 Sztuka. Rozrywki. Sport
  - 8 Językoznawstwo. Nauka o literaturze. Literatura piękna
  - 9 Archeologia. Prehistoria. Geografia. Biografie. Historia<sup>57</sup>.

Wyrazem interesującego nas podejścia kategoryzacyjnego do organizacji UKD są kategorie leksykalne:

- formy dokumentu
- języka
- miejsca
- rasy
- grupy etnicznej i narodowości
- czasu
- punktu widzenia
- sposobu ujęcia tematu
- materiału i osoby.

Z punktu widzenia gramatyki UKD tworzą one kategorię gramatyczną poddziałów wspólnych, czyli poddziałów pomocniczych o nieograniczonym zakresie stosowania, wyrażających pewne ogólne charakterystyki dokumentów poza samym przedmiotem i jego aspektem. Można wyróżnić wśród nich dwie kategorie zbiorcze – kategorię reprezentującą cechy pozat treściowe dokumentów i kategorię reprezentującą cechy treściowe, uznane za relewantne dla reprezentacji i wyszukiwania informacji. W kategorii cech pozat treściowych zgromadzono symbole reprezentujące formę dokumentu i język, w kategorii cech treściowych – symbole reprezentujące: miejsca, rasy, grupy etniczne i narodowości, czas, punkt widzenia, sposób ujęcia tematu, materiał i osobę. Kategoriami odnoszącymi się do własności prezentacji treści są zatem kategoria formy i kategoria języka wyrażane za pomocą symboli poddziałów

---

<sup>55</sup> UKD 1997.

<sup>56</sup> Ścibor 1995.

<sup>57</sup> Szczegółową i wyczerpującą charakterystykę zawartości poszczególnych działów UKD można znaleźć w: Sosińska 1995, s. 101-134.

wspólnych formy i języka. W praktyce celowi temu służą także symbole nie zawarte explicite w odpowiednich tablicach poddziałów wspólnych, ale możliwe do utworzenia w wyniku dołączania symboli poddziałów analitycznych. Symbole poddziałów analitycznych (stanowiące obok symboli syntetycznych jedną z dwóch podgrup symboli poddziałów specjalnych UKD) wyrażają własność lub szczegółowy aspekt badawczy, charakterystyczne dla przedmiotów określonej dziedziny. Na przykład w dziale<sup>58</sup> 81 Językoznawstwo wyróżniono między innymi: -25 Języki potoczne (mówione); -26 Języki pisane. Języki literackie; '271 Poprawność językowa. Błędy językowe; '36 Gramatyka; '367.6 Części mowy; '37 Semantyka; '373 Leksykologia; '373.45 Wyrazy obce. Wyrazy zapożyczone. Kalki językowe itd.

Pod pojęciem formy w UKD rozumie się najczęściej: typ dokumentu, jego cechy fizyczne, cechy sygnału, za pomocą którego utrwalona jest informacja, cechy nośnika informacji, cechy układu informacji w dokumencie, cechy sposobu jej zorganizowania, formę piśmienniczą i wydawniczą, adres czytelniczy<sup>59</sup>. Kategoria leksykalna formy, będąca jednocześnie kategorią gramatyczną o wykładniku (0...) (czytaj: nawias zero), zorganizowana jest hierarchicznie, na ogół według relacji generycznej, na przykład:

- (047) Sprawozdania. Informacje o wydarzeniach. Reportaże
- (047.36) Raporty (biuletyny) handlowe. Raporty marketingowe
- (047.53) Wywiady
- (07) Wydawnictwa dydaktyczne. Podręczniki
- (075) Podręczniki szkolne i akademickie
- (075.2) Podręczniki dla szkół podstawowych. Elementarze
- (075.3) Podręczniki dla szkół średnich
- (075.4) Podręczniki do samokształcenia. Samoucзки
- (075.8) Podręczniki dla szkół wyższych. Skrypty.
- (075.9) Podręczniki dla różnego rodzaju kursów szkoleniowych

Symbole z kategorii **języka**, mające wykładnik gramatyczny = (znak równości), służą do oznaczania języka, w którym dokument został napisany, lub z którego, czy na który został przetłumaczony. Ponadto znalazły się tu dwa symbole o charakterze metainformacyjnym – symbol =00 oznaczający wielo-

---

<sup>58</sup> Wszystkie przykłady w tym rozdziale zostały pobrane z: UKD 1997.

<sup>59</sup> Ten model semantyczny nie jest jednak całkowicie zgodny z prawdą, gdyż wśród symboli wspólnych formy z jakichś powodów znalazł się symbol odnoszący się do cech treści, a nie formy jej przedstawienia. Jest to symbol (091) oznaczający historyczny aspekt przedstawienia przedmiotu dokumentu stosowany w opisie dokumentów prezentujących historię, dzieje dowolnego rodzaju obiektu, z wyjątkiem państw, których historię oznaczają symbole z działu 93/99. W dziale 0 Zagadnienia ogólne i w dziale 80 Językoznawstwo znajdziemy ilustrację sytuacji odwrotnej, tylko bowiem encyklopedie i słowniki poszczególnych dziedzin klasyfikuje się za pomocą symboli poddziałów wspólnych formy. Encyklopedie i słowniki ogólne klasyfikuje się za pomocą symboli z działów 0 lub 80.

języczność i poliglotyzm oraz symbol =03 znakujący teksty tłumaczone i tłumaczenia z języka obcego. Jednostki leksykalne kategorii języka są także uporządkowane hierarchicznie – częściowo według relacji generycznej, w większości jednak według jakiejś bliżej nieokreślonej relacji hierarchizującej. Na przykład symbol =131 oznacza grupę języków italoromańskich, której istnienia nie rejestrują powszechnie używane wydawnictwa informacyjne z zakresu językoznawstwa.

Kategoriemi odnoszącymi się do cech treściowych dokumentów jest siedem pozostałych wcześniej wymienionych kategorii, grupujących symbole należące do poddziałów wspólnych miejsca, rasy, grupy etnicznej i narodowości, czasu, punktu widzenia, sposobu ujęcia tematu, materiału i osoby.

Kategoria leksykalna **miejsca**, zawierająca symbole poddziałów wspólnych o wykładniku (1/9) (nawias okrągły, wewnątrz ciągi cyfr zaczynające się od 1 do 9), służy do lokalizacji przestrzennej treści dokumentu. Dzieli się na cztery fasety: Miejsce i przestrzeń w ogólności, Środowisko fizyczno-geograficzne, określenie miejsca, Świat starożytny według podziału administracyjno-politycznego, Świat nowożytny według podziału administracyjno-politycznego. W obrębie każdej z faset występują klasy pomocnicze obejmujące symbole poddziałów analitycznych. Wyrażają one dodatkowe ogólne charakterystyki przestrzenne, takie jak: strefy określonego miejsca, strony świata, administracyjno-polityczne jednostki terytorialne, podział regionalny, własności wspólnot i związków państw, pole działania inicjatywy prywatnej lub kolektywnej, charakterystyki miejsca z punktu widzenia rozmieszczenia, pochodzenia, przebiegu i przeznaczenia. Leksyka w kategorii miejsca uporządkowana jest hierarchicznie, w większości przypadków według relacji generycznej, jedynie w wypadku państw i miast świata starożytnego i nowożytnego według relacji mereologicznej.

Poddziały wspólne **rasy, narodowości i grupy etnicznej** o wykładniku (=...) (nawias okrągły, znak równości) służą do oznaczania tzw. aspektu etnicznego i narodowościowego treści wyrażonej symbolem głównym oraz grup ludności, których dotyczy dokument. Symbole należące do tej kategorii tworzy się w zasadzie metodą rozbudowy równoległej według wzorca zawartego w poddziałach wspólnych języka (jeśli kryterium wyodrębnienia grupy jest wspólny język) i poddziałach wspólnych miejsca (jeśli kryterium wyodrębnienia grupy jest wspólne terytorium zamieszkania). Należy jednak pamiętać, że w wypadku symboli oznaczających rasy analogie te nie zawsze są całkowite. Nieujawniony *explicite* porządek jednostek leksykalnych reprezentujących tę kategorię można traktować jako odwzorowanie porządku przyjętego w wyżej wymienionych kategoriach dla niej wzorcowych.

Kategoria leksykalna **czasu** o wykładniku "...” (cudzysłów) służy do oznaczania lokalizacji czasowej treści dokumentu. Lokalizacja ta może się wyrażać oznaczeniem daty, pór roku, miesięcy, tygodni, dni, czasu pracy i odpoczynku, relatywnego czasu trwania, częstotliwości, epoki, chronologii

geologicznej, archeologicznej, kulturowej, następstwa chronologicznego, czasu trwania zdarzenia oraz pojęć z zakresu fenomenologii czasu (np. ewolucja, postęp, jednorazowość, powtarzalność) itd. W obrębie tej kategorii określone zostały podkategorie gromadzące jednostki leksykalne przeznaczone do wyrażania dat, jednostki wyrażające czas trwania, periodyczność, chronologię geologiczną, archeologiczną i kulturową oraz pojęcia z zakresu fenomenologii czasu. W poszczególnych podkategoriach leksyka jest uporządkowana hierarchicznie według relacji generycznej.

Organizacja kategorii czasu została wykorzystana w niektórych działach głównych UKD do rozbudowy symboli głównych lub analitycznych. „Na przykład w dziale 7 Sztuka określenie czasu jest związane z określeniem stylu, dlatego wszędzie tam, gdzie wskazanie stylu jednoznacznie wskazuje też okres, w którym dzieło sztuki powstało, nie jest konieczne stosowanie poddziałów wspólnych czasu. Podobnie w dziale 93/99 (od 1994 r. w dziale 94) Historia nie zawsze potrzebne jest korzystanie z tych poddziałów pomocniczych, gdyż określenie czasu bywa tu wyrażone przez symbol główny lub analityczny, np.:

- 940.1 Historia Europy okresu średniowiecza, 374-1492 (do 1994 r.)
- 94(411).01 Historia Szkocji, okres galicki do 1057 (od 1994 r.)
- 940.53 Ogólna historia II wojny światowej 1939-1945 (od 1994 r.)
- 940.55 Historia Europy po 1945 r. (do 1994 r.)”<sup>60</sup>.

W praktyce symbole poddziałów wspólnych czasu mogą pełnić dodatkową funkcję selekcyjną związaną z wielkością i profilem zbiorów, o czym zdają się świadczyć na przykład takie symbole zastosowane w *Przewodniku Bibliograficznym*:

726.04.033.4(438)A/Z:736.033.4.046]”11” – dla publikacji o tytule *Romańskie Drzwi Płockie*;

75.036.2:929-052]”18/19”(03) – dla publikacji o tytule *Encyklopedia impresjonizmu*;

94(438).072”1848”:016 – dla publikacji o tytule *Wiosna Ludów w Wielkopolsce 1848 r.: bibliografia w wyborze*.

Kategoria leksykalna punktu widzenia i sposobu ujęcia tematu grupuje symbole służące do oznaczania tych dwóch aspektów omówienia przedmiotu (przedmiotów) dokumentu. Dzieli się na dwie podkategorie – podkategorię punktu widzenia o wykładniku .00 (kropka, zero, zero) i podkategorię sposobu ujęcia tematu o wykładniku .000 (kropka, zero, zero, zero). „Poddziały wspólne

---

<sup>60</sup> Sosińska 1995, s. 172-173.

punktu widzenia oznaczają tzw. obiektywny punkt widzenia, standardowy aspekt ogólny, z którego rozpatrywana jest większość tematów. Poddziały wspólnego sposobu ujęcia tematu zwane są poddziałami aspektu subiektywnego, czy subiektywnego punktu widzenia, a więc wskazują tzw. punkt widzenia autora związany zwykle z jego światopoglądem, specjalizacją itp.”<sup>61</sup>. Na przykład:

- .001 Punkt widzenia teoretyczny
  - .001.18 Prognozy
  - .001.3 Opisy. Porządkowanie. Porównywanie
  - .001.4 Próby. Doświadczenia. Poszukiwania
  - .001.5 Badania
    - .001.53 Badania laboratoryjne
    - .001.55 Badania polowe
    - .001.57 Badania na modelach
- .002 Punkt widzenia wykonawczy
  - .002.2 Produkcja. Wytwarzanie
  - .002.3 Surowce. Składniki. Elementy
- .003 Punkt widzenia ekonomiczny
- .004 Punkt widzenia użytkowania i funkcjonowania
- .005 Punkt widzenia oprzyrządowania, instalacji i wyposażenia
- .006 Punkt widzenia lokalizacji, przestrzeni i pomieszczeń
- .007 Punkt widzenia personelu (kadr)
- .008 Punkt widzenia organizacyjny

Leksyka w ramach każdej podkategorii kategorii punktu widzenia zorganizowana jest hierarchicznie według relacji generycznej lub hierarchii tematycznej.

.000.0/9 Poddziały wspólnego sposobu ujęcia tematu przez autora

Dalszy podział jak 0/9, np.:

- .000.17 Moralny punkt widzenia
- .000.282 Katolicki punkt widzenia
- 613.88.000.282 Wychowanie seksualne z punktu widzenia religii katolickiej
- 791.86.000.17 Walki zwierząt z moralnego punktu widzenia

Kategoria leksykalna materiału o wykładniku -03 (kreska, zero, trzy) służy do oznaczania surowców lub składników, z których zostały wykonane przedmioty omówione w dokumencie<sup>62</sup>. Rozbudowa symboli opiera się w znacznym

<sup>61</sup> Sosińska 1993, s. 155.

<sup>62</sup> Symboli poddziałów wspólnych materiału używa się wtedy, gdy materiał nie stanowi głównego przedmiotu dokumentu, a jest jedynie charakterystyką uzupełniającą. W wypadku, gdy materiał jest przedmiotem dokumentu, stosuje się odpowiednie symbole główne – najczęściej z działów 66/67.

stopniu na rozbudowie gałęzi symboli głównych zawartych w działach 54/55 i 66/67, czyli gałęzi obejmujących problematykę chemii teoretycznej i przemysłowej oraz złóż mineralnych i ich przetwórstwa. Słownictwo należące do tej kategorii uporządkowane jest według relacji generycznej w siedmiu fasetach:

- „-032 Materiały mineralne występujące w przyrodzie
- 033 Materiały przetworzone pochodzenia mineralnego (gips, wapno, cement, beton, ceramika, kamionka)
- 034 Metale
- 035 Materiały pochodzenia organicznego (drewno, korek, papier, papa, skóry, barwniki, futra, oleje, tłuszcze, woski)
- 036 Tworzywa wielocząsteczkowe. Kauczuk i tworzywa sztuczne
- 037 Tekstyli (włóczka, przędza, tkaniny, sukno)”<sup>63</sup>.

Poddziały wspólne charakterystyk osobowych o wykładniku -05 (kreska, zero, pięć) służą do oznaczania zagadnień dotyczących osób i charakterystyk osobowych<sup>64</sup>. Słownictwo należące do tej kategorii uporządkowane jest w ośmiu fasetach:

- „-051 Osoby jako czynnik aktywny (czynny), podmiot działania (np. nauczający, obsługujący itp.)
- 052 Osoby jako czynnik pasywny (bierny), przedmiot działania (np. obsługiwani, klienci, użytkownicy itp.)
- 053 Osoby według wieku albo grupy wiekowej
- 054 Osoby według charakterystyk etnicznych, narodowościowych albo obywatelskich (np. cudzoziemcy, wygnańcy, repatrianci, obywatele danego kraju itp.)
- 055 Osoby według płci i pokrewieństwa
- 056 Osoby według czynników fizycznych i psychicznych (np. osoby zdrowe fizycznie, inwalidzi wskutek choroby, inwalidzi wskutek wypadku, chorzy umysłowo itp.)
- 057 Osoby według zatrudnienia (np. osoby zatrudnione w pełnym wymiarze czasu pracy, osoby zatrudnione w niepełnym wymiarze czasu pracy, pracownicy fizyczni, robotnicy, pracownicy umysłowi, urzędnicy, personel kierowniczy, pracodawcy, emeryci, uczniowie, studenci itp.)
- 058 Osoby w zależności od sytuacji społecznej i statusu cywilnego (np. bogaci, klasa średnia, nędzarze, osoby niezamężne (nieżonate), małżeństwa, sieroty, dzieci zaadoptowane itp.)”<sup>65</sup>.

---

<sup>63</sup> UKD 1997, s. 38.

<sup>64</sup> Symboli poddziałów wspólnych charakterystyk osobowych również używa się wtedy, gdy są one charakterystykami uzupełniającymi, jeśli punkt widzenia osób jest drugoplanowy w stosunku do przedmiotu dokumentu lub jeśli dla charakterystyk osobowych nie przewidziano w danym dziale symboli głównych lub analitycznych.

<sup>65</sup> Sosińska 1993, s. 159-160.

Na kategoryzację w UKD można także popatrzeć z punktu widzenia kategorii gramatycznych i wyróżnić – kategorię poddziałów wspólnych, rozbitą na dziesięć klas leksykalnych wyżej omówionych, i kategorię symboli głównych, usystematyzowaną na pierwszym stopniu podziału w postaci układu dziesięciu grup dziedzin wiedzy i działalności. Wykładniki morfologiczne przynależności do klasy głównej pełnią jednocześnie funkcję semantyczną. Na przykład: wykładnik 6... stojący na początku każdego symbolu głównego oznacza, że w znaczeniu danego wyrażenia zawarty jest składnik semantyczny ‘nauki stosowane’ lub ‘medycyna’ lub ‘technika’ lub ‘rolnictwo’ (klasa 6 – Nauki stosowane. Medycyna. Technika. Rolnictwo); wykładnik 8 oznacza, że w znaczeniu wyrażenia zawarty jest składnik semantyczny ‘język’ lub ‘lingwistyka’ lub ‘filologia’ lub ‘literatura piękna’ (klasa 8 – Język. Lingwistyka. Filologia. Literatura piękna); wykładnik .000 – element semantyczny ‘bycia punktem widzenia’ itd. Z punktu widzenia wartości składniowej wyrażenia UKD mają wartość albo obligatoryjną autosyntaktycznych wyrażen zdaniotwórczych (symbole główne), albo fakultatywnych wyrażen synsyntaktycznych (symbole poddziałów wspólnych).

Tak więc kategoryzacja leksykalna zastosowana w Uniwersalnej Klasyfikacji Dziesiętej wykazuje cechy wspólne ze schematami stosowanymi w wielu językach informacyjno-wyszukiwawczych, nie tylko typu klasyfikacyjnego. Ustrukturalizowane hierarchicznie kategorie leksykalne Uniwersalnej Klasyfikacji Dziesiętej odwzorowują (na ogół) podstawowe aspekty rzeczywistości, a nie wiedzy, reprezentując cechy semantyczne współtworzące wraz z innymi komponentami znaczenia rozpoznawanymi dzięki relacjom paradygmatycznym konotację symboli UKD. Implikują i równocześnie same są implikowane przez cechy semantyczne współtworzące konotację bezpośrednią i pośrednią poszczególnych jednostek leksykalnych. Z literatury przedmiotu nie wynika, jakie kategorie naturalne i w jaki sposób stały się modelem kategorii w UKD. Nie będziemy też spekulować na ten temat. Chcielibyśmy jednak sformułować jako jeden z wniosków postulat jawnego wiązania modelu kategoryalnego języka informacyjno-wyszukiwawczego z zasadnym i wiarygodnym modelem rzeczywistości, którą język ma opisywać. Symbole główne UKD generalnie reprezentują wymiar Przedmiot (dokumentu), obejmujący znaczenia należące do różnych kategorii ontologicznych. Ponieważ znaczenia te określane są przez wskazanie w pierwszej kolejności dziedziny, do której należą, mamy tu do czynienia z sytuacją, w której ujawnione zostają zarówno własności kategoryalne, jak i dyscyplinarnie określony aspekt rozważań zawartych w klasyfikowanym dokumencie. Na przykład, w dziale 81 Językoznawstwo symbole główne reprezentują poszczególne języki traktowane jako przedmioty badawcze lingwistyki; ich rodzaje identyfikowane są przez wskazanie powiązań etnicznych, np.: 811.111 Język angielski; 811.162.1 Język polski. Do działu tego włączono również języki sztuczne, takie jak języki programowania, które jedynie w wyjątkowych sytuacjach stają się przedmiotem



badan̄ lingwistycznych i s̄ā wówczas reprezentowane przez symbol 811.93. Traktowane jako przedmiot badan̄ z zakresu programowania komputerowego s̄ā reprezentowane przez symbol 519.682. Na koniec chcemy zwrócić uwagę, że chociaŹ podział dyscyplinarny utrwalony w UKD (i wielu innych klasyfikacjach piśmiennictwa) szybko ulega dezaktualizacji, nierazko będąc przeszkodą w odzwierciedlaniu pewnych związków logicznych i tematycznych między klasami i potencjalnymi klasami klasyfikacji, to jednak z punktu widzenia pewnych kryteriów relewancji nadal pozostaje efektywnym narzędziem wyszukiwania informacji. Wydaje się więc, że optymalna organizacja systemu leksykalnego jiw w dokumentacyjnym systemie informacyjno-wyszukiwawczym powinna, obok kategoryzacji, uwzględniać właściwy podział dyscyplinarny uniwersum, zapewniający lokalizację zarówno dyscyplin tradycyjnych, jak i rozwijających się obszarów interdyscyplinarnych i multidyscyplinarnych paradygmatów badawczych.

#### KATEGORIE W JĘZYKACH HASEŁ PRZEDMIOTOWYCH

Model języka haseł przedmiotowych zastosowany w praktyce uzyskał postać różnych jego „odmian”, których cechami wspólnymi pozostały obecnie już jedynie podział słownictwa na dwie kategorie syntaktyczne – tematy, będące wyrażeniami autosyntaktycznymi i określniki, będące wyrażeniami synsyntaktycznymi – oraz pierwszoplanowa rola przedmiotu w opisie dokumentu. Dalsze rozwinięcie kategorii tematu i określnika warunkowane jest przynależnością danego jhp do określonej „odmiany”. Rozważania zawarte w rozdziale dotyczą w zasadzie dwóch języków haseł przedmiotowych – jhp opracowanego przez Adama Łysakowskiego i jhp KABA<sup>66</sup>. Źródłem materiału językowego charakteryzującego pierwszy z wymienionych języków były prace Łysakowskiego<sup>67</sup> i Jadwigi Ćwiekowej<sup>68</sup> oraz *Instrukcja tematowania i katalogu przedmiotowego*<sup>69</sup>. Zarówno w publikacjach Ćwiekowej, jak i w *Instrukcji...* nawiązuje się bowiem, choć nie zawsze explicite, do modelu jhp stworzonego przez Łysakowskiego. Przykłady wyrażen̄ języka haseł przedmiotowych KABA zaczerpnięto z kartoteki wzorcowej jhp KABA oraz z publikacji o tym języku<sup>70</sup>.

---

<sup>66</sup> Język haseł przedmiotowych KABA (Katalogi Automatyczne Bibliotek Akademickich) jest od 1994 r. współtworzony przez dziewięć bibliotek wyższych uczelni (Uniwersytet Warszawski, Uniwersytet Jagielloński, Akademia Górniczo-Hutnicza, Uniwersytet Gdański, Uniwersytet Wrocławski, Katolicki Uniwersytet Lubelski, Uniwersytet Marii Curie-Skłodowskiej, Akademia Rolnicza w Lublinie, Politechnika Lubelska) i wykorzystywany w katalogach online (OPAC) tych bibliotek.

<sup>67</sup> Łysakowski 1946, 1928.

<sup>68</sup> Ćwickowa 1988, 1974.

<sup>69</sup> Sadowska 1991.

<sup>70</sup> KABA 1999, Kartoteka 1994.

Przyjmuje się, że kategoria tematu grupuje wyrażenia wykorzystywane do oznaczania przedmiotów dokumentów lub, w szczególnych przypadkach, ich formy. Praktyka wykazuje, że stwierdzenie to nie zawsze jest prawdziwe. Niejednokrotnie w wielu językach haseł przedmiotowych, żeby oznaczyć pewne przedmioty dokumentu, trzeba użyć, przewidzianego przez projektanta danego jhp i wyrażonego *explicite* w postaci odsyłacza (lub tzw. tropu odsyłacza – jhp KABA) całkowitego, połączenia tematu z określnikiem. Wprawdzie przypadki takie nie są liczne, ale jednak się zdarzają. Jeśli dla uproszczenia wyeliminujemy z rozważań tego typu wyjątki, to możemy przyjąć, że dla języka haseł przedmiotowych charakterystyczne są dwie kategorie tematów: kategoria tematów oznaczających desygnaty bezpośrednie (dokumenty) i kategoria tematów oznaczających desygnaty pośrednie (to, o czym mowa w dokumentach).

Wśród autorów zajmujących się problematyką języków haseł przedmiotowych panuje zgoda co do tego, że tematy rzeczowe, czyli należące do kategorii tematów oznaczających desygnaty pośrednie, mogą być wyrażane albo za pomocą nazw pospolitych (tematy ogólne), albo nazw własnych lub deskrypcji pełniących funkcję nazw własnych (tematy jednostkowe, indywidualne). Łysakowski wśród tematów jednostkowych<sup>71</sup> umieścił tematy osobowe, geograficzne i topograficzne oraz tematy, które nazwał jednostkowymi rzeczowymi, oznaczające „a) wydarzenia historyczne, b) instytucje i organizacje, c) dzieła anonimowe, d) ewent. inne jednostki rzeczowe”<sup>72</sup>, nie precyzując, jakie tematy można i należy zaliczyć do „innych jednostek rzeczowych”. Ćwiekowa w klasie tematów, które nazwała indywidualnymi (osobowe, językowo-etniczne, etniczno-terytorialne, geograficzne, chronologiczne, rzeczowe), także wyróżniła podklasę tematów rzeczowych i określiła je następująco: „tematy rzeczowe, poza nazwami instytucji i organizacji, mogą odzwierciedlać nazwy własne różnych przedmiotów i pojęć, np. czasopism, pierwiastków chemicznych itp.”<sup>73</sup>. Podobnie w *Instrukcji tematowania i katalogu przedmiotowego*<sup>74</sup> wśród tematów jednostkowych<sup>75</sup> wyróżniono tematy rzeczowe jednostkowe, czyli „a) tytuły dzieł zbiorowych, anonimowych, wydawnictw ciągłych (...); b) nazwy systemów informacyjnych, komputerowych, języków programowania itp. (...); c) nazwy urządzeń technicznych, pojazdów (jeśli zdecydowaliśmy się te przedmioty wprowadzić w postaci samodzielnych tematów)”<sup>76</sup>. Także w języku KABA przyjęto, że z punktu

---

<sup>71</sup> Tematy jednostkowe „są to nazwy własne przedmiotów niepowtarzalnych, oznaczonych indywidualnie, jako jednostki historyczne lub geograficzne”. Łysakowski 1946, s. 76.

<sup>72</sup> Łysakowski 1946, s. 168.

<sup>73</sup> Ćwickowa 1988, s. 82.

<sup>74</sup> Sadowska 1991, s. 58.

<sup>75</sup> „Tematy jednostkowe reprezentowane są przez nazwy jednostkowe, tj. takie, które mają jeden desygnat. Najczęściej wyrażane są za pomocą imion własnych lub tzw. deskrypcji jednostkowych, tj. grup wyrazów używanych do identyfikacji jakiegoś obiektu, indywidualnego przedmiotu”. Sadowska 1991, s. 21.

<sup>76</sup> *Ibidem* s. 22.

widzenia formy reprezentacji „jednostki leksykalne JHP mogą być wyrażane przez nazwy ogólne i nazwy jednostkowe”<sup>77</sup>. Hasła wyrażane nazwami jednostkowymi podzielono dalej na osobowe, korporatywne, geograficzne i tytułowe (tytuły ujednolicone).

Zastanawia, skąd wzięła się tendencja zaliczania do tematów jednostkowych „innych jednostek rzeczowych”, z których część to tematy pozornie jednostkowe o mnogich desygnatach pośrednich. Być może z jednej strony pewną rolę odegrał tu sposób nazywania ich desygnatów sugerujący, że mamy do czynienia z nazwą własną, z drugiej zaś to, że przynajmniej część z nich odnosi się zwykle do całej serii identycznych z założenia przedmiotów (mnożość jednorodna). Jeśli wyłączymy ewidentnie błędnie zakwalifikowane przedmioty, na przykład nazwy pierwiastków chemicznych (temat rzeczowy indywidualny według Ćwiekowej), to możemy przyjąć, że pozostałe tematy rzeczowe jednostkowe powinny być albo ideonimami (tytuły dzieł zbiorowych, anonimowych i wydawnictw ciągłych), albo chrematonimami<sup>78</sup>. W katalogach i indeksach przedmiotowych do najczęściej spotykanych obiektów identyfikowanych za pomocą chrematonimów należą: środki transportu (np.: statki morskie, rzeczne, samoloty, pasażerskie, samochody); niektóre rodzaje broni i obiektów wojskowych (np. okręty wojenne, samoloty bojowe, działa, miecze); sprzęt gospodarstwa domowego, sprzęt audiowizualny, sprzęt i oprogramowanie komputerowe; dzwony; rzadkie okazy natury, które wtórnie traktowane są jako obiekty kultury materialnej.

Bez szkody dla praktyki i z korzyścią dla teorii można zatem przyjąć, że w ogólnym modelu języka haseł przedmiotowych, biorąc jako kryterium rodzaj desygnatów, tematy (i jak dalej wykażemy również określniki) można podzielić na: tematy rzeczowe, tj odwzorowujące cechy desygnatów pośrednich wyrażen jhp, i tematy formalne, odwzorowujące cechy desygnatów bezpośrednich. Biorąc z kolei pod uwagę sposób wyrażania tematu, wyróżnia się wśród nich tematy o desygnatach jednostkowych (tematy jednostkowe) i tematy o desygnatach mnogich (tematy ogólne). Tematy ogólne mają postać wyrażen pospolitych nadawanych w języku naturalnym zbiorom przedmiotów jednakowych i powtarzalnych. Mogą to być tematy o charakterze klasowym, formalnym lub przedmiotowym, desygnujące obiekty o cechach nie indywidualnych, np.: **Medycyna, Psychologia, Pisarki, Szkoły wyższe, Antologie, Ssaki,**

---

<sup>77</sup> Kartotcka 1994, s. 34.

<sup>78</sup> „Chrematonimy (gr. *chrema*, *chrematos* = rzecz, przedmiot towar) to nazwy własne niektórych materialnych wytworów ręki ludzkiej, przemysłowych lub rękodzielniczych, jednostkowych lub seryjnych, takich, które nie są trwale związane z określonym krajobrazem. Zakres użycia tego terminu nie jest jeszcze w pełni ustalony. Czasami niesłusznie (nieczgodnie z etymologią terminu) zalicza się do chrematonimów tzw. idconimy (np. tytuły utworów literackich, plastycznych i muzycznych, filmów, czasopism i dzienników, nazwy organizacji politycznych i społecznych), a także nazwy własne innych obiektów (np. kin, teatrów, will, domów wczasowych, lokali gastronomicznych i handlowych, zakładów przemysłowych, jednostek wojskowych, zespołów muzycznych)”. Kosyl 1993, s. 439.

**Spółdzielnie, Czasopisma polskie.** Tematy ogólne zazwyczaj nie podlegają dalszemu podziałom. Jedynie w *Instrukcji...* wyodrębniono kategorię leksykalną tematów ogólnych językowo-etnicznych charakteryzującą się tym, że wszystkie należące doń tematy posiadają w zestawie cech znaczeniowych własność ‘bycia w określonym języku’ lub ‘przynależności do określonej grupy etnicznej czy państwa’. *Instrukcja...* nie ogranicza zakresu łączliwości cechy językowo-etnicznej, zwracając jedynie uwagę, że „warta jest podkreślenia przy tematach związanych z wytworami literatury, sztuki i kultury”, zaś inne „tematy językowo-etniczne należy wprowadzać ostrożnie”<sup>79</sup>.

W zasadzie można przyjąć, że cytowani wyżej autorzy<sup>80</sup>, różnie nazywający klasę tematów jednostkowych<sup>81</sup>, są zgodni co do dalszego jej podziału na tematy: osobowe, geograficzne i toponimiczne, chronologiczne, korporatywne, etniczne, chrematonimiczne i ideonimiczne (dzieł zbiorowych, anonimów, wydawnictw ciągłych)<sup>82</sup>. Ćwickowa wyodrębniła także tematy językowo-etniczne i etniczno-terytorialne, będące nazwami „poszczególnych języków (i ich grup językowych) lub innych pojęć językoznawczych oraz zjawisk związanych z przynależnością do jakiegoś narodu bądź jego kultury przejawiających się w więzi języka lub terytorium”<sup>83</sup>. Potrzebę istnienia tej podkategorii tematów jednostkowych uzasadniała następująco: „Tematy językowo-etniczne zostały wyróżnione przeze mnie w grupie tematów indywidualnych ze względu na rozwój piśmiennictwa, potrzebę wydzielenia literatury związanej cechami etnicznymi spośród piśmiennictwa omawiającego ogólnie język, literaturę, sztukę i inne zjawiska związane z kulturą oraz ze względu na uznawanie tematów wielowyrazowych za właściwą postać wyrażania pojęć czy zjawisk złożonych. Łysakowski wyróżniał nazwy języków jako określniki językowe, łącząc je z określnikami gatunkowymi. Helena Hleb-Koszańska wyróżniała w funkcji określników gatunkowych – obok oznaczania gatunku przedmiotu – pojęcia etniczno-kulturalne wyrażane <<przydawką geograficzną>>, np.: Język polski, Koń arabski, Tatarzy krymscy (...). Dlatego też do grupy tematów indywidualnych można zaliczyć tematy językowo-etniczne”<sup>84</sup>. Przypomnijmy w tym miejscu myśl Tadeusza Kotarbińskiego „nazwa – naród francuski – chociaż zbiorowa jest jednostkowa zarazem, bo jeden jest tylko przedmiot, o którym można orzec prawdziwie, iż jest ogółem Francuzów”<sup>85</sup>. W *Instrukcji*

<sup>79</sup> Sadowska 1991, s. 49.

<sup>80</sup> Wyjątkiem jest tu język KABA, w którym nie wszystkie wymienione kategorie zostały wyróżnione explicite, co nie znaczy, że nie występują w nim tematy wszystkich wyróżnionych kategorii. KABA jest językiem, którego wykładnia teoretyczna dopiero powstaje i tu raczej tkwi przyczyna jego odrębności, nie zaś w jakiejś szczególnej odmienności strukturalno-funkcjonalnej.

<sup>81</sup> Ćwickowa używa terminu *temat indywidualny* nie *jednostkowy*.

<sup>82</sup> Użyte określenia osobowy, korporatywny, chrematonimy itd. zostały przyjęte na użytek pracy w celu ujednoczenia nazewnictwa stosowanego przez poszczególnych autorów. Terminologia właściwa różnym autorom jest używana w załączonych cytatach.

<sup>83</sup> Ćwickowa 1988, s. 81.

<sup>84</sup> Ibidem.

<sup>85</sup> Kotarbiński 1961, s. 22.

tematowania i katalogu przedmiotowego wyróżniono kategorię tematów językowo-etnicznych, ale jako wchodzącą w skład kategorii tematów ogólnych, co z punktu widzenia natury tematów językowo-etnicznych wydaje się być jedynym uzasadnionym rozwiązaniem. Naszym zdaniem Ćwiekowa niefortunnie połączyła problem nacechowania językowo-etnicznego tematu z problemem tematów oznaczających grupy etniczne. Za *Instrukcją...* przyjmiemy, że podkategorię w obrębie kategorii tematów jednostkowych tworzą tematy etniczne, mając świadomość, że wśród językoznawców nie ma powszechnej zgody co do przynależności etnonimów, czyli nazw narodów, narodowości, plemion i mniejszych grup etnicznych oraz pojedynczych reprezentantów tych grup do kategorii nazw własnych.

Krytyczna analiza cytowanych wyżej tekstów prowadzi do wniosku, że ze względu na cechy desygnatów pośrednich wśród tematów języka haseł przedmiotowych można wyróżnić następujące kategorie leksykalne:

- językowo-etniczne
- chronologiczne
- etniczne
- geograficzne i toponimiczne
- korporatywne
- osobowe
- ideonimiczne i chrematonimiczne<sup>86</sup>.

Obecnie przyjrzymy się kategoriom tematów wyróżnianych ze względu na rodzaj oznaczanych cech dokumentu. Są to tematy oznaczające cechy treściowe (rzeczowe) dokumentów (tematy przedmiotowe i klasowe) oraz tematy oznaczające cechy formalne. Łysakowski tak określał te trzy kategorie tematów. „Tematem (czysto) przedmiotowym jest nazwa oznaczonego przedmiotu dzieła. Mogą to być wszelkiego rodzaju przedmioty: realne i idealne; konkretne całości, usamodzielnione części i oderwane cechy (własności); przedmioty fizyczne i psychiczne; osoby i rzeczy; zjawiska i dyspozycje; czynności i stany; jakości i stosunki; fakty i zagadnienia; przedmioty należące do minionej rzeczywistości, jako też przedmioty dopiero projektowane (jeszcze nie istniejące). (...) Tematami klasowymi są nazwy klas czyli dziedzin życia kulturalnego i gospodarczego (...) [zaś] tematem formalnym jest nazwa rodzaju piśmienniczego lub wydawniczego, nazwa gatunku tekstów, utworów, dzieł, wydawnictw, książek, druków”<sup>87</sup>. Podobnie typologię tematów według kryterium rodzaju oznaczanych cech dokumentów przeprowadza Ćwiekowa<sup>88</sup>. W *Instrukcji...* można znaleźć jedynie definicję tematu formalnego, którym jest

---

<sup>86</sup> Tematy językowo etniczne zaliczane są do szerszej kategorii tematów ogólnych, podczas gdy tematy chronologiczne, etniczne, geograficzne i toponimiczne, korporatywne, osobowe, ideonimiczne i chrematonimiczne nazywane są tematami jednostkowymi (indywidualnymi).

<sup>87</sup> Łysakowski 1946, s. 75-76, 81.

<sup>88</sup> Ćwickowa 1988, s. 78-80.

„temat odzwierciedlający formę piśmienniczą lub wydawniczą dokumentu”<sup>89</sup>. W języku KABA z punktu widzenia referencji „wyrażającej się w określeniu obszaru odniesienia danej jednostki leksykalnej w stosunku do uniwersum cech charakteryzujących dokumenty”<sup>90</sup> tematy dzielone są na rzeczowe (treściowe) i tematy formy. Dopelnieniem kategorii tematów formalnych powinna być kategoria tematów nie będących formalnymi, czyli treściowych (rzeczowych). Ten typ opozycji formalny ÷ nieformalny ma istotne znaczenie dla praktyki posługiwania się językami haseł przedmiotowych. Podział tematów nieformalnych na przedmiotowe i klasowe jest ciekawym problemem teoretycznym jhp, choć jeszcze dość słabo wyjaśnionym przez zwolenników tej koncepcji podziału tematów języków haseł przedmiotowych. Ze względu na rodzaj oznaczanych cech dokumentów kategorię tematów w jhp można zatem podzielić na dwie podkategorie: tematów formalnych i tematów nieformalnych, które bywają nazywane tematami rzeczowymi lub treściowymi.

Kategoria określników grupuje synsyntaktyczne wyrażenia jhp przeznaczone do charakteryzowania treści lub formy dokumentów. Dla polskiej tradycji języka haseł przedmiotowych charakterystyczny był podział określników na trzy następujące grupy:

„a) Określniki szczegółowe, odmienne dla różnych tematów:

1. jednostkowe – podają indywidualne szczegóły tematów jednostkowych (...):  
Słowacki Juliusz – „Król Duch”  
Lwów – Ossolineum  
Grunwald – bitwa 1410 r.
2. gatunkowe włącznie z językowymi (...):  
Katalog przedmiotowy  
Powieść polska
- b) Określniki ogólne, schematyczne, wspólne dla tematów wszystkich rodzajów, wskazują ujęcie lub formę dzieła (...):
3. klasowe i inne rzeczowe (...):  
Wartość – filozofia  
Malarstwo – wystawy
4. formalne (...):  
Radio – czasopisma
- c) Określniki lokalizujące, także wspólne dla różnych tematów, wskazują umiejscowienie tematu w przestrzeni i czasie (...):
5. geograficzne (...):  
Teatr – Francja
6. chronologiczne (...):  
Polska – XX w.”<sup>91</sup>.

<sup>89</sup> Sadowska 1991, s.13.

<sup>90</sup> Karoteka 1994, s. 34.

<sup>91</sup> Łysakowski 1946, s. 101-102.

Zastanawia wyodrębnienie określników gatunkowych rozumianych jako „stałe gatunki przedmiotu, gałęzie dziedziny (...), odmiany formy piśmiennej”<sup>92</sup>. Analiza określników zakwalifikowanych przez Łysakowskiego jako gatunkowe w kontekście wyszczególniającej funkcji tematów prowadzi do wniosku, że tzw. tematy z określnikami gatunkowymi należałoby raczej traktować jako tematy wieloelementowe wyodrębnione relacją podrzędności generycznej (gatunek) z istniejącego w danym jhp lub potencjalnego tematu nadrzędnego (rodzaj). Zauważmy, że jakkolwiek układ określników prowadzi u Łysakowskiego do innego niż w wypadku tematów schematu ich uporządkowania, to w istocie odwołuje się do tych samych kryteriów podziału<sup>93</sup>.

Ćwiewkowa przeprowadziła podział określników bardzo podobny do zaproponowanego przez Łysakowskiego. Wśród określników szczegółowych wyodrębniła określniki „specjalne (określające swoiste, właściwe cechy każdego tematu lub grup tematów), np.: Pieniądze – obieg; Produkcja – jakość; Przestępstwo – zapobieganie. (...) Określniki specjalne są charakterystyczne prawie dla każdego tematu. Rola ich wzrasta coraz bardziej, gdyż opracowuje się je w ścisłej korelacji z przedmiotem badań naukowych lub potrzeb społecznych. Opracowanie określników specjalnych wymaga od redagującego katalog przedmiotowy doskonałej orientacji w merytorycznej istocie danego problemu. Użytkownik natomiast dzięki określnikom specjalnym uzyskuje literaturę dotyczącą danego tematu opracowaną katalogowo pod względem metodologicznym poprawnie, zgodnie z właściwościami danego zagadnienia. Przykłady określników specjalnych: Organizacja Narodów Zjednoczonych – członkostwo; Organizacja Narodów Zjednoczonych – organizacje wyspecjalizowane; Rysunek – techniki”<sup>94</sup>. Autorka sugeruje opracowanie zestawów określników specjalnie zaprojektowanych dla poszczególnych tematów lub „grup tematów”. Zauważmy jednak, że propozycja ta w niczym nie narusza przedstawionego wcześniej podziału kategorialnego określników według kryteriów rodzaju desygnatów pośrednich i rodzaju oznaczanych cech dokumentów, co najwyżej dodaje jeszcze jeden możliwy podział określników na określniki właściwe dla jednego tematu, określniki właściwe dla pewnych grup tematów, aż po określniki właściwe dla wszystkich tematów.

Rozwinięcie tej myśli znajdziemy w języku KABA, który wprowadza do kategorii określników kolejny element różnicujący. Jest to podział określników na kategorię tzw. określników swobodnych, czyli takich, które mogą być łączone z dowolnymi tematami i kategorię tzw. określników związanych, czyli takich, które mogą być łączone tylko z tym tematem lub tematami, z którymi połączył je projektant języka, na przykład określniki związane

---

<sup>92</sup> Łysakowski 1946, s. 107.

<sup>93</sup> Z podziału zostały wyłączone określniki gatunkowe, które są w istocie szczególnego typu tematami, zaś w proponowanym podziale, ze względu na ich ogromne zróżnicowanie, mogłyby się znaleźć w każdej z kategorii.

<sup>94</sup> Ćwiewkowa 1988, s. 85-87.

chronologiczne właściwe dla danego kraju. Pomędzy tymi dwiema kategoriami lokuje się trzecia, o charakterze mieszanym, czyli kategoria tzw. określników typowych. Określniki typowe w KABA mają większy zakres łączliwości od określników związanych, ale nie są w pełni swobodne. Na obecnym etapie w języku KABA wyróżniono następujące klasy określników typowych stosowanych po nazwach: kategorii osób, określających grupy wiekowe, społeczne, polityczne, zawodowe, ideowe, artystyczne, wyznaniowe itp.; kategorii pisarzy; literatur i gatunków literackich; języków i grup języków; instytucji, organizacji, kongresów, konferencji, sympozjów, wystaw, targów, festiwalu itp.; szkół i uczelni; typów szkół i uczelni; sił zbrojnych; wojen; gałęzi gospodarki; typów czasopism; grup etnicznych; form muzycznych; instrumentów muzycznych; osobowych; obiektów geograficznych i fizjograficznych, w tym obszarów wodnych; parlamentów; zgromadzeń zakonnych, w tym zakonów ścisłych, kongregacji, bractw, instytutów religijnych, wspólnot i ruchów; religii i wyznań; obiektów o konotacji religijnej; przedmiotów o konotacji prawnej; po tytułach utworów autorskich ksiąg świętych; po tytułach utworów anonimowych; po tytułach ujednoczonych; po nazwach organizmów żywych, np.: zwierząt, roślin; narządów wewnętrznych i części ciała; chorób; produktów i tworzyw; produktów chemicznych; pojazdów naziemnych.

Podsumowując, można powiedzieć, że charakterystyka kategoriałna języków haseł przedmiotowych ma naturę wieloaspektową. Biorąc pod uwagę rodzaj desygnatu, można wyróżnić dwie kategorie leksykalne – kategorię wyrażen (tematów i określników) oznaczających desygnaty bezpośrednie jednostek leksykalnych jhp i kategorię wyrażen oznaczających desygnaty pośrednie<sup>95</sup>. Każdej z zawartych w nich jednostek przysługuje ponadto cecha przynależenia do jednej z dwu kategorii – przedmiotów konkretnych (tematy i określniki jednostkowe, indywidualne) lub przedmiotów niekonkretnych (tematy i określniki ogólne). Do kategorii wyrażen oznaczających desygnaty bezpośrednie należą tematy i określniki formalne (formy). Kategorie leksykalne wyróżniane w kategorii wyrażen oznaczających desygnaty pośrednie są różne w różnych odmianach jhp. Dla scharakteryzowanych języków haseł przedmiotowych Adama Łysakowskiego i jhp KABA właściwe są następujące kategorie leksykalne: osoby; obiektu geograficznego i fizjograficznego; wydarzenia historycznego; instytucji, organizacji i tzw. imprezy; grupy etnicznej; tytułu. Dodatkowo należałoby wyróżnić kategorię leksykalno-gramatyczną wyrażen chrematonimicznych, której cechą dystynktywną jest specyficzna forma wyrażen – wszystkie zawierają dopowiedzenie<sup>96</sup>, na przykład: **Bismarck (pancernik)**, **Tadeusz Kościuszko (samolot)**, **Wisła (telewizor)**, **Kreolki (wafle)** – wskazujące denotaty tych wyrażen.

<sup>95</sup> Ungurian nazwałby te dwie kategorie raczej wymiarami: Przedmiotu i Treści oraz Formy.

<sup>96</sup> Tematy innych kategorii leksykalnych też mogą zawierać dopowiedzenia. Ale w przypadku chrematonimów dopowiedzenie jest elementem obowiązkowym każdego tematu.



Na kategoryzację w jhp można także popatrzeć z punktu widzenia kategorii syntaktycznych i wyróżnić dwie z nich – kategorię tematów i kategorię określników. Tematy tworzą kategorię zdaniotwórczych obligatoryjnych wyrażeń autosyntaktycznych, określniki zaś kategorię niezdaniotwórczych synsyntaktycznych wyrażeń fakultatywnych. Synsyntaktyczne określniki tworzą z kolei subkategorie syntaktyczne ze względu na różny stopień ich łączliwości z tematami.

W przeszłości języki haseł przedmiotowych powstawały z myślą o odwzorowywaniu rzeczywistości konkretnego zbioru dokumentów i sytuacji konkretnej biblioteki. Były językami o charakterze aposteriorycznym w przeciwieństwie do wielu apriorycznych systemów klasyfikacyjnych. Konsekwencją tego podejścia była nieprzejmowalność (lub co najwyżej ograniczona przejmowalność) języków haseł przedmiotowych. Obecnie obserwujemy zjawisko przeciwne. W warunkach zautomatyzowanych systemów informacyjno-wyszukiwawczych w subjęzyku wyszukiwawczym przestała też bezwzględnie obowiązywać zasada gramatyczna dotycząca struktury hasła przedmiotowego i odwzorowywanego przez tę strukturę pewnego porządku gramatyczno-semantycznego. Nie jest zatem pozbawione sensu pytanie, co pozostało niezbywalną cechą wszystkich obecnie istniejących języków haseł przedmiotowych. Naszym zdaniem jest to przedmiotowe podejście do tekstu oraz respektowanie w subjęzyku informacyjnym podziału kategoryalnego na autosyntaktyczne tematy i synsyntaktyczne określniki.

Tradycyjnie w językach haseł przedmiotowych, w przeciwieństwie do języków typu klasyfikacyjnego i języków deskryptorowych, nie porządkowano całej substancji znaczeniowej języka, tj. nie wyodrębniano wymiarów, superkategorii, kategorii czy klas głównych, podkategorii, faset lub klas i podklas, aż do najniższego poziomu pojedynczych wyrażeń. Poprzestawano na łączeniu jednostek leksykalnych relacjami nadrzędności i podrzędności, doprowadzającymi do powstania mniej lub bardziej rozbudowanych drzew hierarchicznych, których najwyższe poziomy hierarchii nie tworzyły spójnego podziału uniwersum leksyki języka, pozostając konglomeratem. Obecnie obserwuje się tendencję do nadawania słownikom jhp pewnych cech tezaurusów, co, jeśli założyć, że celem jest przejęcie wszystkich pozytywnych cech tezaurusów, wiązałoby się z jawną prezentacją całości słownictwa jhp jako systemu leksykalno-semantycznego. Do tej pory nieliczne ze znanych języków haseł przedmiotowych zrealizowały (z różnym efektem) postulat dostosowania słownika jhp do modelu tezaurusu. Być może dla wielu języków haseł przedmiotowych jest to postulat nierealny ze względu na często uniwersalny zakres jhp i liczebność słownictwa. Na przykład jhp RAMEAU, opracowany i wykorzystywany między innymi w Bibliothèque nationale de France, w 1999 roku liczył około 220 tys.<sup>97</sup> tematów i określników. Semantyczne uporządkowanie

---

<sup>97</sup> Dla porównania przypomnimy, że aktywny słownik, na przykład studenta, przeciętnie liczy 12-14 tys. wyrazów ogólnych.

takiej liczby wyrażen wydaje się być zadaniem przekraczającym możliwości nie tylko bibliotekarzy. Innym aspektem tego problemu jest szansa odnalezienia w tak licznyim ustrukturalizowanym formalnie (abecedłowo) zbiorze wyrażenia reprezentującego poszukiwane znaczenie i, co za tym idzie, możliwości pełnego wykorzystywania potencjału semantycznego tego słownika, na który składają się uszeregowane abecedłowo w sześciu tomach tysiące drzewek hierarchicznych. Z punktu widzenia użytkownika języka prawdopodobieństwo użycia w zapytaniu informacyjnym któregoś z wyrażen tego języka (lub jego ekwiwalentów) wydaje się być dość małe, zwłaszcza w odniesieniu do wyrażen tzw. szczegółowych.

Zagadnienie optymalnej wielkości słownika jiw w różnych systemach informacyjno-wyszukiwawczych, w tym liczby jednostek leksykalnych, wydaje się być ważnym problemem poznawczym o znaczących konsekwencjach dla efektywności posługiwania się językami informacyjno-wyszukiwawczymi w procesach tworzenia reprezentacji dokumentów i wyszukiwania informacji. Generalnie można przyjąć, że wraz ze wzrostem liczby jednostek leksykalnych rośnie znaczenie właściwej struktury semantycznej języka i jej prezentacji. Naszym zdaniem, płaska struktura pola semantycznego może nie mieć wpływu na efektywność tylko w przypadku niewielkich słowników (możliwych do opanowania pamięciowego) języków paranaturalnych, w których znaczenia jednostek leksykalnych nie różnią się istotnie od znaczeń równokształtnych wyrażen języka naturalnego. W językach o strukturze płaskiej lub mającej przewagę cech struktury płaskiej ciężar identyfikacji relewantnych jednostek leksykalnych spoczywa na użytkowniku i bibliotekarzu opracowującym zbiory. Oznacza to, że na przykład użytkownikowi, który nie znalazł szukanej jednostki leksykalnej ani wśród wyrażen danego języka, ani wśród wyrażen odrzuconych (wówczas może się spodziewać odsyłacza kierującego do wyrażenia przyjętego w jiw) pozostaje żmudne i, nie dające gwarancji powodzenia, „zgadywanie” sposobów, za pomocą których można wyrazić interesujące go znaczenie. Większość języków informacyjno-wyszukiwawczych, poza wspomnianymi wcześniej językami o nielicznym i paranaturalnym słownictwie, wymaga takiego ustrukturalizowania słownika, żeby wyszukiwanie informacji (w słowniku i za pomocą jiw w zbiorze informacyjnym systemu) było efektywne. W pewnych przypadkach użytkownikowi wystarczy ujawnienie w słowniku podziału wiedzy naukowej i wiedzy o działalności praktycznej człowieka, w innych właściwsza byłaby relewantna kategoryzacja elementów rzeczywistości, którą język opisuje. Najlepsze rezultaty powinno dawać połączenie trzech obecnie wykorzystywanych form strukturalizacji pola semantycznego, tj. podziału dyscyplinarnego, kategoryalnego i struktur relacyjnych. Podział dyscyplinarny i podział kategoryalny wyznaczają pewien makroporządek informacji odwzorowanej w strukturze słownika jiw i w zbiorze informacyjno-wyszukiwawczym systemu, podczas gdy organizację pojęć reprezentowanych przez jednostki leksykalne wewnątrz klas wyodrębnionych przez dyscypliny lub kategorie określają, znajdujące odwzorowanie

w paradygmatyce jiw, różnego rodzaju relacje semantyczne tworzące strukturę relacyjną jiw. Dodatkowo efektywność posługiwania się językiem informacyjno-wyszukiwawczym mogą podnosić „przyjazne” użytkownikowi sposoby i formy jego udostępniania i prezentacji w systemie oraz komunikacji z samym siw.

#### KATEGORIE W TEZAURUSIE TECHNOLOGII ŻYWNOSCI I PRZEMYSŁU SPOŻYWCZEGO<sup>98</sup>

Wybrany tezaurus jest słownikiem języka deskryptorowego z dziedziny technologii żywności i przemysłu spożywczego. Składa się z części alfabetyczno-hierarchicznej, zawierającej 13 281 artykułów deskryptorowych i 5362 artykuły askryptorowe, części systematycznej oraz trzech kategorii identyfikatorów<sup>99</sup> (nazwy firmowe preparatów enzymatycznych, nazwy firmowe urządzeń, nazwy serów). Pozostałe nazwy własne uzyskały status deskryptorów, ponieważ albo liczyły zbyt mało jednostek leksykalnych, żeby, zdaniem twórców *Tezaurusu*, utworzyć z nich odrębny wykaz, albo dlatego, że są równokształtne z nazwami typów obiektów. Na przykład, deskryptory KADŹ WHIRPOOL. i DYFUZOR DDS. są nazwami firmowymi urządzeń i zarazem wskazują określony typ konstrukcji kadzi i dyfuzorów.

Pełen artykuł deskryptorowy składa się z następujących elementów:

- deskryptor<sup>100</sup> tytułowy;
- określenie lub definicja (w nawiasach kwadratowych), np. HOMOZYGOTA. [organizm powstały z połączenia gamet o jednakowym składzie genetycznym]; LETALNOŚĆ. [wyrażona w procentach śmiertelność populacji na jednostkę czasu];
- kody adresowe deskryptora tytułowego w części systematycznej tezaurusu;
- askryptory poprzedzone wskaźnikiem relacji ekwiwalencji Nu (Nie używaj);
- deskryptory bezpośrednio szersze (wskaźnik relacji SD) i deskryptory bezpośrednio węższe (WD) w stosunku do deskryptora tytułowego;
- deskryptory kojarzeniowe (KD), reprezentujące pojęcia często współwystępujące w dokumentach z pojęciem reprezentowanym przez deskryptor tytułowy;

<sup>98</sup> Tezaurus 1996.

<sup>99</sup> „Identyfikator – wyrażenie aut syntaktyczne języka deskryptorowego równokształtne z nazwą własną”. Słownik 1993a, s. 46. W praktyce identyfikatory nie zawsze są nazwami własnymi sensu stricto; zawsze nazywają jednak obiekty z założenia identyczne (tzw. mnogość jednorodna).

<sup>100</sup> Wykładnikiem formalnym deskryptora jest ”.” (kropka) jako ostatni jego znak graficzny i dodatkowo, przynajmniej w dostępnej wersji drukowanej tezaurusu, zapis dużymi literami.

– nazwa samodzielnej kategorii (nazwy samodzielnych kategorii) w części systematycznej, w której występuje deskryptor tytułowy, poprzedzona wskaźnikiem GK.

Przykład artykułu deskryptorowego:

SOK SUROWY.

20.195.018.012.

20.246.006.

21.333.051.015.

21.810.051.

21.819.

21.894.015.

Nu moszcz

świeży sok

SD SOK CUKROWNICZY.

SOKI.

SOK SYROPOWY.

ZALEWA.

ŻYWNOSĆ MINIMALNIE PRZETWORZONA.

WD SOK DYFUZYJNY.

SOK UZYSKANY MECHANICZNIE.

KD CZĘŚCI ROŚLIN.

EKSTRAKCJA.

EKSTRAKTOR.

KRAJANKA BURACZANA.

ODCIEKACZ.

OWOCE I WARZYWA.

PRASA.

SYROP SKROBIOWY.

SYROP ZIEMNIACZANY.

TŁOCZENIE.

WIRÓWKA.

GK PRZETWORY CUKROWNICZE.

PRZETWÓR SYROPIARSKI.

SOKI.

ZALEWA.

ŻYWNOSĆ MINIMALNIE PRZETWORZONA.

W *Tezaurusie technologii żywności i przemysłu spożywczego* wyróżniono 20, jak to nazwano, „podstawowych kategorii semantycznych deskryptorów”. Nazwy niektórych z nich są jednocześnie deskryptorami. Określając makroporządek tezausa, niektóre kategorie podzielono na kategorie składowe. Otrzymana struktura przedstawia się następująco:

- A. Cechy formalne dokumentu
- B. Dziedziny i dyscypliny nauki i techniki, gałęzie gospodarki narodowej oraz sfery działalności praktycznej człowieka

|         |   |
|---------|---|
| B I.    | DZIAŁY I GAŁĘZIE GOSPODARKI NARODOWEJ.    |
| B II.   | DZIEDZINY I DYSCYPLINY NAUKI I TECHNIKI.  |
| B III.  | SFERY DZIAŁALNOŚCI PRAKTYCZNEJ CZŁOWIEKA. |
| C.      | CIAŁA FIZYCZNE.                           |
| D.      | Czas                                      |
| E.      | ENERGIA.                                  |
| F.      | IMPREZA.                                  |
| G.      | METODY <sup>101</sup> .                   |
| H.      | Miejsce                                   |
| I.      | ORGANIZACJE I INSTYTUCJE.                 |
| J.      | ORGANIZMY.                                |
| K.      | Osoby                                     |
| L.      | PROCESY.                                  |
| M.      | Przedmioty                                |
| M I.    | CZĘŚCI I WYDZIELINY ORGANIZMÓW.           |
| M II.   | MATERIAŁY OPAKOWANIOWE I KONSTRUKCYJNE.   |
| M III.  | OBIEKT BUDOWLANY.                         |
| M IV.   | ODPADY.                                   |
| M V.    | OPAKOWANIA.                               |
| M VI.   | PREPARATY.                                |
| M VII.  | PRODUKTY ŻYWNOŚCIOWE.                     |
| M VIII. | PRZEPISY.                                 |
| M IX.   | URZĄDZENIA.                               |
| N.      | RODZAJ ZASTOSOWANIA.                      |
| P.      | SUBSTANCJE CHEMICZNE.                     |
| R.      | System                                    |
| S.      | Teorie                                    |
| T.      | WŁASNOŚCI.                                |
| U.      | Wydarzenie                                |
| W.      | ŹRÓDŁA I NOŚNIKI ENERGII.                 |

Z tego, co napisano we Wstępie do części systematycznej *Tezaurusu*, wynika, że podział kategorii B i M był spowodowany jedynie względami technicznymi („Ze względów czysto praktycznych (brak odpowiedniej liczby liter w alfabecie) (...) wyodrębniono dwie grupy kategorii oznaczane literami B i M, a należące do nich kategorie oznaczono daną literą i kolejnymi cyframi rzymskimi”<sup>102</sup>). Do kategorii WŁASNOŚCI. włączono deskryptory określające stany i relacje, np.: CHOROBA., STOSUNEK SKŁADNIKÓW., ZANIECZYSZ-

<sup>101</sup> Przydzielając deskryptory do kategorii PROCESY. i METODY. przyjęto, że określeniem procesu jest termin mówiący o wyniku działania, np. fermentacja, korozja, wycłaczanie, zamrażanie; terminy mówiące o tym jak lub za pomocą czego dany wynik jest uzyskiwany, np.: zamrażanie fluidyzacyjne, kriogeniczne, owiewowe, powietrzne zostały zakwalifikowane do kategorii METODY.

<sup>102</sup> Tezaurus 1996, s. IX.

CZENIE., traktowane jako wskaźniki jakości obiektów materialnych lub procesów i metod. Stąd na przykład w artykule deskryptora CHOROBA. (zawierającym askryptory *stan chorobowy* i *stan choroby*) jako deskryptor bezpośrednio szerszy (SD) występuje WADA. a nie STAN.

Kategorie podstawowe podzielono na 1309 „samodzielnych kategorii deskryptorów”. Deskryptory należące do poszczególnych kategorii zorganizowano według zasad klasyfikacji fasetowej. Niektóre deskryptory występują w więcej niż jednej subkategorii tej samej lub różnych kategoriach podstawowych. Na przykład, deskryptor BETANINA. został wprowadzony do subkategorii GLIKOZYDY. i POCHODNE FENYLOALANINY. w kategorii P. SUBSTANCJE CHEMICZNE. oraz do subkategorii PREPARAT POCHODZENIA ROŚLINNEGO. w kategorii MVI PREPARATY. Dzięki temu rozwiązaniu stworzono możliwość lepszego rozróżniania znaczeń, a także zwiększono liczbę tzw. punktów dostępu do deskryptora. W części alfabetyczno-hierarchicznej *Tezaurusu* zjawisko eksponowania relewantnych dla użytkownika stałych własności obiektów manifestuje się w postaci więcej niż jednego deskryptora szerszego i/lub więcej niż jednej kategorii głównej w artykule deskryptorowym. Jedyną kategorią, której elementy nie mogą migrować, jest RODZAJ ZASTOSOWANIA. Zgrupowano w niej nazwy sytuacyjnych własności obiektów, np.: DODATEK DO ŻYWNOŚCI., PÓLPRODUKT., ZAMIENNIK. Umieszczenie któregoś z tych deskryptorów w innej kategorii sprawiłoby, że podział pola semantycznego byłby oparty częściowo na relacjach stałych (rodzaj-gatunek), a częściowo na relacjach sytuacyjnych, co z kolei mogłoby być przyczyną błędów w wyszukiwaniu.

- 21.006. CHLEB
  - zob. kat. PIECZYWO PIEKARSKIE. 21.285.051.
  - (wg gatunku)
- 21.006.003 – CHLEB BŁONWIT.
  - [chleb o podwyższonej zawartości błonnika]
- 21.006.006 – CHLEB CHRUPKI.
  - zob. kat. PIECZYWO TRWAŁE. 21.288.003.003.
- 21.006.009. – CHLEB GLUTENOWY.
- 21.006.012. – CHLEB GRAHAM.
- 21.006.015. – CHLEB KONSERWOWY.
  - [chleb wypiekany i sterylizowany w puszkach, o trwałości do kilkunastu miesięcy]
- 21.006.018. – CHLEB LECYTYNOWY.
  - [chleb z dodatkiem lecytyny]
- 21.006.021. – CHLEB SKROBIOWY.
- 21.006.024. – CHLEB TOSTOWY.
- 21.006.027. – PUMPERNIKIEL.
  - [chleb z mąki razowej]

- (wg kształtu)
- 21.006.030. – CHLEB OKRĄGLY.  
 21.006.033. – CHLEB PODŁUŻNY.  
 –
- (wg sposobu rozrostu)
- 21.006.036. – CHLEB DESKOWY.  
 21.006.039. – CHLEB KOSZYCZKOWY.  
 –
- (wg wyciągu mąki)
- 21.006.048. – CHLEB CIEMNY.  
 21.006.051. – CHLEB JASNY.  
 21.006.054. – CHLEB PEŁNOZIARNISTY.  
 21.006.057. – CHLEB PYTLÓWY.  
 21.006.060. – CHLEB RAZOWY.  
 21.006.063. – CHLEB SITKOWY.

Analiza struktury kategorialnej *Tezaurusu technologii żywności i przemysłu spożywczego* prowadzi do dwóch ogólnych wniosków. W języku z zakresu technologii żywności i przemysłu spożywczego, w przeciwieństwie do wcześniej omawianych języków informacyjno-wyszukiwawczych o zakresie uniwersalnym, termin kategoria używany jest w sposób wskazujący na bliskie pokrewieństwo znaczeniowe z terminem kategoria w pracach CRG (ogólna klasa znaczeniowa wyodrębniona na podstawie cech dziedzinowych, branżowych). Zastosowana kategoryzacja jest przykładem kategoryzacji specjalistycznej, mającej zaspokoić potrzeby organizacji pojęć w określonym obszarze wiedzy i dla określonych użytkowników. Wyodrębnione kategorie, z których pewne odpowiadają zdefiniowanym przez Unguriana wymiarom<sup>103</sup>, dotyczą zarówno własności samych dokumentów, jak i omówionych w nich przedmiotów i treści. Zwraca uwagę obszernie i wiarygodnie uzasadnione pragmatyczne i dokumentacyjne podejście projektantów języka deskryptorowego z dziedziny technologii żywności i przemysłu spożywczego do zagadnienia kategoryzacji i organizacji języka w ogóle<sup>104</sup>. Zamiast dzielić pole

<sup>103</sup> Zauważmy, że kategoria A odpowiada wymiarowi cech formalnych, kategoria B – wymiarowi treści, kategoria D – wymiarowi czasu, kategoria H – wymiarowi przestrzeni, pozostałe kategorie – wymiarowi przedmiotu.

<sup>104</sup> Uzasadnieniem tak sformułowanej opinii niech będzie cytat ze Wstępu do części alfabetycznej *Tezaurusu*, wyjaśniający dlaczego zrezygnowano z wykorzystania modyfikatorów. „W trakcie tworzenia języka zrezygnowano z pierwotnego, zgodnego z życzeniem zleceniodawcy, zamiaru uwzględnienia w jego słownictwie niewielkiego zestawu modyfikatorów. Modyfikatorami są zazwyczaj terminy ogólne o dużym stopniu łączliwości (np. badania, otrzymywanie, prognozowanie, organizowanie, zastosowanie), które, po ich połączeniu za pomocą określonego symbolu z odpowiednimi deksyptorami lub identyfikatorami (np. SERY. + badania, SESJA. + organizowanie, SOK JABŁKOWY. + badania, ŻEL. + badania, SOK JABŁKOWY. + otrzymywanie, PRODUKCJA. + prognozowanie, PRASA SHENKA. + badania) precyzują w charakterystyce aspekty treści dokumentu lub zapytania informacyjnego, w celu zapobieżenia wyszukiwaniu informacji nierelevantnych do zapytań. W ramach kontroli kompletności słownictwa podstawowego JTŻiPS zaindeksowano (...) ponad 2 tysiące dokumentów. Wówczas okazało się, że z większości

znaczeniowe języka na wymiary i ewentualnie kategorie ogólne, te zaś z kolei na specjalistyczne, uniwersum leksyki języka podzielono od razu na kategorie specjalistyczne. Model kategorialny omawianego języka wynika ze struktury rzeczywistości, którą ten język opisuje.

W języku z dziedziny technologii żywności i przemysłu spożywczego została wyróżniona tylko jedna kategoria gramatyczna deskryptorów z formalnym wykładnikiem w postaci kropki, będących jednocześnie elementami autosyntaktycznej kategorii obligatoryjnych wyrażen zdaniotwórczych. Wykładniki morfologiczne przynależności do określonej kategorii leksykalnej, tj. cyfry i liczby do 30, pełnią jednocześnie funkcję semantyczną. Na przykład: wykładnik 22 stojący na początku każdego kodu deskryptora oznacza, że dany deskryptor lokuje się w kategorii 22 (M VIII – PRZEPISY., przy czym sam deskryptor PRZEPISY. ma kod 22.003.); wykładnik 01 oznacza, że dany deskryptor oznacza cechy formalne dokumentu, takie jak jego język i typ (kategoria A – Cechy formalne dokumentu).

Podsumowując rozważanie struktury kategorialnej trzech wybranych języków informacyjno-wyszukiwawczych, należy stwierdzić, że cechy semantyczne wyróżniane w poszczególnych językach informacyjno-wyszukiwawczych w postaci układu kategorii można podzielić na te, które wynikają z cech właściwych wszystkim (lub większości) jiw i te, które są ściśle związane z zakresem denotacyjno-konotacyjnym języka. Cechami semantycznymi, które charakteryzują większość języków informacyjno-wyszukiwawczych, są komponenty znaczenia wynikające z funkcji metainformacyjnej jiw. Zwykle dają się sprowadzić do trzech następujących własności: 'być o pewnym przedmiocie (treści dokumentu)', 'być w pewnym aspekcie', 'być w pewnej formie (mieć pewne własności prezentacji treści)', choć nie zawsze własności te są tak nazwane. W wielu językach informacyjno-wyszukiwawczych można wskazać klasy cech semantycznych reprezentujących różne szczegółowe przypadki ogólnych własności 'bycia o pewnym przedmiocie', 'bycia w pewnym aspekcie' i 'bycia w pewnej formie'. Bywa również i tak, że własności te, zwłaszcza 'bycia o pewnym przedmiocie' i 'bycia w pewnym aspekcie', są kumulowane na osi leksykalnej, a rozdzielane na osi syntagmatycznej języka. Wyróżnione trzy grupy cech semantycznych tworzących konotację wyrażen jiw dodatkowo świadczą o wielowymiarowej naturze języków informacyjno-wyszukiwawczych, o czym była mowa na s. 140-143.

---

terminów, które miały być modyfikatorami, należy utworzyć deskryptory ze względu na występowanie w literaturze publikacji o tematyce bardzo ogólnej (np. na temat finansowania badań dotyczących żywności lub nt. metod prognozowania wyników procesów technologicznych). Tczaurus 1996, s. XVI.



Organizację zasobów informacyjnych w sieci Internet, dostępnych poprzez przeglądarki<sup>105</sup> WWW, scharakteryzujemy na przykładzie wyszukiwarki AltaVista. Pozostałe wyszukiwarki wprowadzić mogą dysponować innymi kategoriami i narzędziami wyszukiwawczymi, ale istota mechanizmu funkcjonowania kategorii i strukturalizacji zbiorów wyszukiwawczych jest analogiczna.

AltaVista dysponuje narzędziami, za pomocą których można wyszukiwać różne kategorie informacji i metainformacji. Wyszukując, można dokonywać selekcji również według kryterium subkodu języka, w jakim ma być szukana informacja, na przykład `image:nazwa_pliku` (wyszukuje pliki graficzne o podanej nazwie). Podstawową jednostką wyszukiwaną jest dokument HTML, zwany też stroną WWW (lub stroną). Informacje mogą być wyszukiwane:

- z właściwego tekstu dokumentu (zwanego BODY) zgodnie z powszechnie znaną zasadą przeszukiwania tekstu swobodnego (z pominięciem tzw. tagów HTML, linków i adresów URL) przy ewentualnym wykorzystaniu operatorów algebry Boole'a, operatorów kontekstowych<sup>106</sup> i maskowania. AltaVista daje również możliwość wyszukiwania z uwzględnieniem w zapytaniu dużych i małych liter, co także ma wpływ na trafność (dokładność) odpowiedzi;
- z części dokumentu zwanej nagłówkiem (HEAD).

Nagłówek strony zawiera wiele różnych kategorii informacji<sup>107</sup>, między innymi:

sposób zakodowania polskich znaków diakrytycznych (<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=...">)  
 język (<META NAME="language" CONTENT="...">)  
 autor (<META NAME="copyright" CONTENT="...">)  
 tytuł (<META NAME="title" CONTENT="...">)  
 zawartość strony (<META NAME="description" CONTENT="...">)  
 słowa kluczowe (<META NAME="keywords" CONTENT="...">).

<sup>105</sup> *Przeglądarka* to program zainstalowany w komputerze użytkownika (w stacji roboczej) i prezentujący na monitorze dokumenty HTML; *wyszukiwarka* zaś to system komputerowy identyfikujący w Internecie serwery ze stronami www, indeksujący je i przechowujący adresy poindeksowanych stron.

<sup>106</sup> Z operatorów kontekstowych najczęściej wykorzystywanymi są " " (cudzysłów), który powoduje, że wyszukiwane zostaną dokumenty, w których występuje cała objęta cudzysłowem fraza a nie jej elementy oraz operator NEAR, powodujący wyszukanie dokumentów zawierających przynajmniej 10 wyrazów, w których zawiera się zadany ciąg znaków.

<sup>107</sup> Wszystkie informacje umieszczane są w nawiasach trójkątnych i poprzedzane wyrazem META.

Podany niżej tekst jest przykładem dokumentu w formacie http (kodzie źródłowym)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML>
<HEAD>
  <META HTTP-EQUIV="Content-Type" CONTENT="text/html;
charset=iso-8859-2">
  <META NAME="Author" CONTENT="Auto Net Sp. z o.o.">
  <META NAME="Description" CONTENT="komis, serwis, dealer
DAEWOO, tico, nexia, espero, lanos, nubira, leganza, polonez, atu, caro,
samochody amerykańskie, jeep gradn cherokee, chrysler">
  <TITLE>Rocar Auto sp. z o.o. – strona główna</TITLE>
</HEAD>
```

Trzy ostatnie kategorie informacji są wyświetlane na monitorze przeszukiwarki w odpowiedzi na zapytanie użytkownika, z tym że słowa kluczowe są prezentowane w kontekście fragmentu właściwego tekstu.

Tak wygląda inny fragment tego samego dokumentu (URL <http://www.rocar.com.pl/>) w formacie edycyjnym

**Rocar Auto Sp. z o. o.**  
**ul. Wyścigowa 58**  
**53-012 Wrocław**  
**tel. komis (071) 362 88 95**  
**tel. salon (071) 339 78 95**  
**tel. serwis (071) 339 79 58**  
**tel. sekretariat (071) 339 79 21**  
**fax (071) 339 79 21**

Poniższe strony najlepiej oglądać w rozdzielczości 800x600

Rocar Auto jest firmą działającą na rynku motoryzacyjnym od 1993 roku. Swoim zasięgiem obejmuje teren Dolnego Śląska, ale jeżeli chodzi o sprzedaż samochodów z rynku amerykańskiego operujemy na terenie całego kraju. Zastanawiając się na sensie istnienia naszej firmy doszliśmy do stworzenia jej misji. Oto ona: Dążymy do doskonałości w świadczeniu usług motoryzacyjnych. Zapewniamy kompleksową, rzetelną i profesjonalną obsługę poprzez nieustanne badanie potrzeb naszych klientów. Jesteśmy dumni z tego co robimy.

Bardzo ważnym elementem strony WWW są słowa kluczowe, będące podstawą indeksowania. Serwery wyszukujące indeksują zasoby sieci na podstawie zawartości sekcji „keywords”. Jeśli twórca strony nie przyporządkował jej słów kluczowych lub w ogóle nie sformułował nagłówka, który nie jest obowiązkowym elementem dokumentu, strona jest indeksowana na podstawie tekstu zawartego w części BODY ze wszystkimi mankamentami indeksowania i wyszukiwania za pomocą swobodnych wyrażen derywowanych z tekstu języka naturalnego (ang. *free-text seraching*).

Słowa kluczowe są też podstawą porządkowania wyszukanych stron. Na ogół ich liczba jest na tyle duża, że użytkownik nie jest w stanie zapoznać się ze wszystkimi. Wyszukiwarka, układając strony w określonej kolejności, w pewnym stopniu wpływa na decyzje użytkownika. Jest bowiem prawdopodobne, że nie mogąc zapoznać się ze wszystkimi, użytkownik wybierze te, które są na początku listy rankingowej. Obserwując zjawisko układania list rankingowych przez wyszukiwarki, można zauważyć, że każdej stronie jest przypisywana jedna i ta sama wartość, którą przez analogię można nazwać wagą. Wartość wagi jest następnie dzielona przez liczbę słów kluczowych przypisanych stronie przez jej twórcę lub ustanowionych na mocy indeksowania swobodnego sekcji BODY. Wynika stąd, że im większa jest liczba słów kluczowych, tym mniejsza ich waga, co w konsekwencji powoduje dalszą pozycję danej strony na liście rankingowej. Relewantność używanych słów i tekstów samych dokumentów HTML ma decydujące znaczenie z punktu widzenia efektywności wyszukiwania. Innym elementem mającym niekwestionowany wpływ na jakość wyszukiwania, są stosowane systemy kategorii. Zdarza się, że autorzy dokumentów przyporządkowują im wiele słów kluczowych o wątpliwej relewancji i niskiej (ze względu na liczbę słów) wadze wyszukiwawczej. Oto przykład takiego dokumentu.

#### **ARE ANIMAL ODORS, GERMS & BACTERIA IN YOUR ENVIRONMENT?**

Well, grab a cup of coffee and be my guest as you discover  
that O<sub>3</sub> is the obvious solution to the problem.  
<http://www.odatus.com>

This page can be found by typing in any of the following search words:

akc, AKC, akc registered, akc registered puppies, akc kennels, akc breeders, stud services, breeding services, exotic animals, veterinarians, veterinarian associations, pet diseases, pet care, pet grooming, pet sitters, pet control, pet stores, feed stores, pet supplies, pet foods, cat breeding, dog breeding, horse breeding, types of cats, types of dogs, types of horses, animal care, Abyssinian cats, American Curl cats, American Shorthair cats, American Wirehair cats, Balinese cats, Birman cats, Bombay cats, British horthair cats, Burmese cats, Chartreux cats, Colorpoint Shorthair cats, Cornish Rex cats, Devon Rex cats, Egyptian Mau cats, European Burmese cats, Exotic cats, exotic cats, Havana Brown cats, Javanese cats, Japanese Bobtail cats, Bobcats, Korat cats, Main Coon cats, Manx cats, Norwegian Forest cats, Ocicat, Oriental cats, Persian cats, Russian Blue cats, Ragdoll cats, Scottish Fold cats, Selkirk Rex cats, Siamese cats, Singapura cats, Somali cats, Tonkinese cats, Turkish Angora cats, Turkish Van cats, American Water Spaniels, Braque Francais, Brittany Spaniels, Clumber Spaniels, American Cocker Spaniels, English Cocker Spaniels, Curly Coated Retrievers, English Setters, English Springer Spaniels, Field Spaniels, Flat Coated Retrievers, French Spaniels, German Long Haired Pointers, German Shorthaired Pointers, German Wirehaired Pointers, Golden Retrievers, Gordon Setters, Irish Setters, Irish Water Spaniels, Labrador Retrievers, Nova Scotia Duck Tolling Retrievers, Pointers, Pudelpointers, Sussex Spaniels, Smooth Haired Vizslas, Wire-haired Vizslas, Weimaraner, Welsh Springer Spaniel, Wirehaired Pointing Griffon, Afghan Hound, Basenji, Bassett Hound, Beagle, Coonhounds, Bloodhound, Borzoi, Dachshunds, Drovers, English Foxhounds, Greyhounds, Harriers, Ibizan Hounds, Irish Wolfhounds, Norrbottenspets, Norwegian Elkhounds, Otter Hounds, Petit Basset

Griffon Vendecens, Pharaoh Hounds, Rhodesian Ridgebacks, Salukis, Scottish Deerhounds, Whippets, Akitas, Alaskan Malamutes, Bernese Mountain Dogs, Boxers, Bullmastiffs, Canaan Dogs, Canadian Eskimo Dogs, Doberman Pinschers, Entlebucher Mountain Dogs, Giant Schnauzers, Great Danes, Great Pyrenees, Greenland Dogs, Karelian Bear Dogs, Komondors, Kuvasz, Leonbergers, Mastiffs, Newfoundlands, Portuguese Water Dogs, otweilers, Saint Bernards, Samoyeds, Schnauzers, Siberian Huskies, Airedale Terriers, American Staffordshires, Australian Terriers, Bedlington Terriers, Border Terriers, Bull Terriers, Cairn Terriers, Dandie Dinmont Terriers, Fox Terriers, Irish Terriers, Kerry Blue Terriers, Lakeland Terriers, Manchester Terriers, Norfolk Terriers, Norwich Terriers, Miniature Schnauzers, Scottish Terriers, Scalyham Terriers, Skye Terriers, Soft Coated Wheaten Terrier, Staffordshire Bull Terriers, Welsh Terriers, West Highland White Terrier, Affenpinschers, Brussels Griffons, Cavalier King Charles Spaniels, Chihuahuas, Chinese Crested Dogs, English Toy Spaniels, Italian Greyhounds, Japanese Chins, Malteses, Manchester Toy Terriers, Mexican Hairless Dogs, Miniature Pinscher, Min Pins, Papillons, Pekingese, Pomeranians, Pugs, Shih Tzus, Silky Terriers, Toy Poodles, Yorkshire Terriers, American Eskimo Dogs, Bichon Frises, Boston Terriers, Bulldogs, Chinese Shar Peis, Chow Chows, Dalmatians, Finnish Spitz, Japanese Spitz, Keeshonds, Lhasa Apsos, Lowchens, Poodles, Schipperkes, Schipperkes, Shiba Inu, Tibetan Spaniels, Tibetan Terriers, Australian Cattle Dogs, Australian Shepherd Dogs, Bearded Collies, Belgian Lackenoi, Belgian Malinois, Belgian Sheepdogs, Belgian Tervurens, Berger Picards, Berger des Pyrenees, Bouvier des Flandres, Briards, Collies, Rough Coat Collies, Smooth Coat Collies, Welsh Corgis, German Shepherds, Norwegian Buhunds, Old English Sheepdogs, Pulis, Shetland Sheepdogs, Shelties, Swedish Vallhunds, Akhal Tekes, Alter Reals, American Shetlands, Andalusians, Anglo Arab, Appaloosas, Arabian Horses, Aricgeois, Asian Wild Horses, Asses, Assateagues, Asturcons, Australian Ponies, Australian Stock Horses, Avlignese, Aztecas, Barbs, Bashkirs, Basuto Ponies, Bavarian Warmbloods, Belgian Warmbloods, Brabants, Breton Drafts, Brumby, Budyonics, Budyonny, Burma Ponies, Camargues, Canadian Cutting Horses, Caspian Horses, Cheval de Selle Francais, Chincotfagufs, Cleveland Bays, Clydesdales, Cob Horses, Colorado Rangers, Connemaras, Criollos, Dale's Ponies, Danish Warmbloods, Dartmoors, Dole Gudbrandsdals, Donkeys, Dutch Warmbloods, East Friesians, Eguus Caballus, Exmoors, Falabellas, Fell Ponies, Franches Montagnes, Frederiksborgs, French Trotters, Friesians, Furiosos, Galicenos, Galloway Ponies, Gelderlands, Groningens, Hackney Horses, Hackney Ponies, Hafingers, Hanoverians, Highland Ponies, Holsteiners, Huculs, Icelandic Horses, Icelandic Ponies, Irish Drafts, Irish Half Breeds, Italian Heavy Draft, Javas, Jutlands, Kabardins, Karabakhs, Kathiawaris, Knabstrups, Koniks, Landais, Lipizzancers, Lusitanos, Manades, Mangalargas, Manipurs, Marcmana, Missouri Fox Trotters, Missouri Fox Trotters, Mongolian Wild Horses, Morgans, Murakosis, Murgese, Mustangs, New Forest Ponies, Norikers, North Swedish Horses, Old English Blacks, Oldenburgs, Orlov Trotters, Paint Horses, Palominos, Paso Finos, Paso Peruvians, Percherons, Persian Arabs, Pintos, Pony Francais de Selle, Pony of the Americas, American Ponies, Quarter Horses, Rhinelanders, Rocky Mountain Ponies, Russian Trotters, Sable Island Ponies, Saddlebreds, Salernos, Sardinians, Shagya Arabs, Shetland Ponies, Shires, Skogruss, Skyos, Sorraias, Standardbred Horses, Standard Bred Horses, Suffolk Punches, Swedish Ardennes, Swedish Warmbloods, Tennessee Walking Horses, Tennessee Walkers, Tersky, Thoroughbreds, Trakehners, Walers, Welsh Mountain Ponies, Wurtenburgs, Zebbras, rabbits, piglets, sheep and goats, sheep, goats, strich.

Na podstawie słów kluczowych lub innych wyrażeń użytych w tekstach dokumentów, wyszukiwarki przypisują dokumenty do kategorii organizujących informacje i dostęp do nich. Na podstawie przeglądu pewnej (być może niereprezentatywnej, ale bardzo trudno jest tu określić kryteria reprezentatywności) liczby dokumentów można stwierdzić, że w wielu przypadkach słowa

kluczowe są dobierane tak, aby jak największa liczba osób „odwiedziła” daną stronę; kwestia relewantności słów jest drugoplanowa, a w niektórych wypadkach wręcz marginalna.

Jako przykład weźmy układ kategorii zaczynający się od kategorii **Entertainment&Media**. Składają się nań kategorie: **Best of the Web; Authors; Bestsellers&Reviews; Bookstores; Chat&Events; Countries&cultures; Genres; Magazines&Journals; Online Texts; Poetry&Plays; Theory&Criticism; Writing&Publishing**. Zauważmy, że kategorie te nie mają nic wspólnego z kategoriami klasycznymi. Być może są one kategoriami kognitywnymi. Z powodu braku informacji trudno o tym wyrokować, a nie wydaje się być uzasadnione nazywanie kognitywnym wszystkiego tego, co nie spełnia warunków klasycznej kategoryzacji. *Kognitywny* nie powinien być traktowany jako synonim *dowolny*. Można jedynie przypuszczać, że omawiane kategorie mają charakter użytkowy i że będą się zmieniać wraz z rozwojem zasobów.

Do dalszej analizy wybraliśmy kategorię **Genres**, którą z kolei tworzą kategorie: **Children; Feminist; Folklore&Myth; Gay&Lesbian; Horror; Hypertext; Literary Fiction; Mystery&Crime; Nonfiction; SF&Fantasy; Romance; Storytelling**. Obecność tych właśnie kategorii i brak innych (wydawałoby się oczywistych) jest prawdopodobnie efektem zastosowania wspomnianej wcześniej indukcyjnej i użytkowej metody tworzenia kategorii. Można się spodziewać, że duża liczba dokumentów na pewien temat spowoduje utworzenie kolejnej kategorii. Na przykładzie tym widać również, jak kategorie wzajemnie na siebie nachodzą (w ujęciu klasycznym kategoria **Literary Fiction** nie mogłaby pozostawać w relacji współrzędności do takich kategorii jak na przykład **SF&Fantasy** czy **Romance**) i przenikają się ze względu na różnorodność kryteriów, według których zostały wyodrębnione (kategoria **Literary Fiction** może i zawiera dokumenty, które jednocześnie znajdują się w kategorii **Children**). Na kolejnym poziomie podziału kategoryalnego w **Literary Fiction** zostały wyróżnione: **African/Middle East; Asian&Indian; Australian; British&Irish; European&Russian; Latin American** i **North American**. W kategorii **European&Russian** można było z kolei znaleźć 57 odesłań (tzw. linków) do stron poświęconych pisarzom (np.: Andersen, Hans Christian; Bulgakov, Mikhail; Goethe, Johann Wolfgang; Milosz, Czesław; Levi, Primo; Simenon, George; Solzhenitsyn, Aleksandr), poetom (choć jest odrębna kategoria **Poetry&Plays**) i literaturom (np.: **Albanian Literature; Dutch Literature, Project Runeberg** – elektroniczna edycja klasyki literatury i sztuki Nordyckiej, **Turkish Poetry**) wraz z krótkimi tekstami charakteryzującymi te strony.

Niejasne, myślące czy przenośne mogą też być nazwy samych kategorii. Na przykład nazwa kategorii **Bibliomania**, w której „herpetolodzy mogą obejrzeć bogatą ofertę książek na temat pielęgnacji, biologii i ekologii, plus znaczki i monety” nie tylko wydaje się być przykładem nadużycia terminu **bibliomania**,

ale wyjęta z kontekstu kategorii nadrzędnych (Hobbies&Interests; Pets&Animals; Unusual Pet Animals; Magazines&Books) mogłaby być całkowicie myląca.

W przyszłości (być może niedalekiej) można się spodziewać, że pewne zasoby informacyjne w sieciach komputerowych (takich jak Internet) zostaną poddane „kontrolom jakości” w jakiejś formie bliskiej, przy całej odmienności środków i techniki, tradycyjnemu, w najlepszym tego słowa sensie, opracowaniu treściowemu informacji. Świadczą o tym między innymi coraz liczniejsze publiczne dyskusje na temat różnych aspektów zarządzania informacją (ang. *information management*) dostępną za pośrednictwem Internetu. Nie można jednak nie widzieć cech pozytywnych organizacji informacji w Internecie. Kategorie dostępne za pośrednictwem wyszukiwarek internetowych wydają się być kategoriami niezwykle specjalistycznymi w sensie związku ze specyfiką organizowanej informacji. Należy się jednak zastrzec, że sformułowane tu wnioski z powodu braku odpowiedniej literatury i ze względu na rozmiary systemu informacyjno-wyszukiwawczego, do jakiego dostęp uzyskuje użytkownik na przykład AltaVisty, należy traktować jako prawdopodobne, a nie pewne. Ktoś mógłby uważać, że twórcy kategorii internetowych projektują je tak, żeby poszczególne kategorie i ich układy miały jak najbardziej naturalny charakter, jak najbliższy sposobowi myślenia i wyszukiwania informacji przez użytkowników Internetu. Nieograniczony dostęp do Internetu sprawia jednak, że owa naturalność musi być dość umownie rozumiana. Trudno bowiem wyobrazić sobie obraz świata „naturalny” dla ludzi wszystkich kultur, religii, poglądów politycznych, etycznych itd. Zatem nie w naturalności czy „poprawności” tkwi moc wyszukiwawcza kategorii internetowych, ale w gotowości internautów do akceptowania proponowanych kategorii, jeśli można za ich pomocą odnaleźć potrzebną informację.

Dla projektantów języków informacyjno-wyszukiwawczych i osób zajmujących się teorią jiw wynikałyby z tego co najmniej dwa, warte dalszych badań, wnioski. Należałoby sprawdzić, czy w przypadku każdego języka informacyjno-wyszukiwawczego budowanego metodą indukcyjną niezbędne jest weryfikowanie jego systemu leksykalno-semantycznego metodą dedukcyjną. Czy zawsze rezultatem będzie optymalizacja języka, czy może w pewnych przypadkach będzie to, zapewne pracochłonna i kosztowna, rozbudowa języka nie mająca wpływu na efektywność posługiwania się nim? Druga refleksja dotyczy struktury systemu kategorii. Zauważmy, że nierozłączność i nieostrość pewnych kategorii oraz różnorodność kryteriów wyodrębniania kategorii tego samego poziomu nie jest postrzegana przez użytkowników Internetu jako wada systemu. Wynika stąd, że należałoby zbadać, jak przy projektowaniu pewnych języków, na przykład przeznaczonych do stosowania w systemach odwzorowujących nie naukowy, a potoczny obraz świata, sprawdziłaby się organizacja struktury semantycznej jiw oparta w większym stopniu na kategoriach kognitywnych niż ontologicznych. Powtarzanie pewnych informacji (bardziej, mniej i bardzo mało typowych) w wielu różnych kategoriach zdaje się być obecnie jednym z elementów siły wyszukiwawczej Internetu. Można zatem

powiedzieć, że Internet w jakimś stopniu jakość zastępuje ilością. Nie zachęcając projektantów jiw do rezygnacji z troski o jakość jiw, uważamy jednak, że powinni zwracać większą uwagę na tworzenie wielu (idealnie: wszystkich) punktów dostępu do kategoryzowanej informacji.

## TWORZENIE KATEGORII W JĘZYKACH INFORMACYJNO-WYSZUKIWAWCZYCH

Patrząc na projektowanie języka informacyjno-wyszukiwawczego w ujęciu modelowym, można wyróżnić cztery główne fazy tego procesu. Są to: określanie zakresu i tworzenie modelu systemu semantyczno-leksykalnego jiw; wybór terminów, opracowanie (w tym strukturalizacja) zgromadzonej terminologii, prezentacja słownictwa. Dokładniej omówimy fazy pierwszą i trzecią, w nich bowiem może się pojawić problem kategoryzacji jako sposobu i metody organizacji pola semantycznego jiw.

Jednym z pierwszych zadań stojących przed projektantem języka informacyjno-wyszukiwawczego jest określenie granic systemu leksykalnego i konstytuujących go pól znaczeniowych<sup>108</sup>, które będą następnie strukturalizowane i wypełniane materiałem językowym. W językowym interpretowaniu świata dużą rolę odgrywa relacja podobieństwa. Podobieństwo jest również jednym z czynników kształtujących pola znaczeniowe. U podstaw teorii pól legło systemowe widzenie leksyki języka, tj. odrzucenie zatimizowanego sposobu widzenia słowa, jego znaczenia i zmian zachodzących w rozwoju języka. Omawiając wybrane zagadnienia pól, mające związek z teorią języków informacyjno-wyszukiwawczych, a zwłaszcza kategoryzacją w tych językach, przyjmujemy pewne uproszczenie, polegające na łącznym rozważaniu planów treści i wyrażania jiw tam, gdzie nie ma to skutków negatywnych. W rezultacie będziemy wymiennie posługiwać się terminami pole semantyczne (znaczeniowe) i pole leksykalne (wyrazowe, znakowe). Nie jest to zresztą przypadek odosobniony, czego jednym z dowodów może być definicja pola semantycznego, mówiąca, że jest to „uporządkowany wewnętrznie pod względem paradygmatycznym i syntagmatycznym zbiór leksemów reprezentujących zbiór pojęć związanych jakimś pojęciem nadrzędnym”<sup>109</sup>.

Przyjmuje się, że słownictwo języka stanowi system<sup>110</sup> wzajemnie ze sobą powiązanych leksemów. Stwierdzenie to, akceptowane we współczesnej semantyce leksykalnej, wywodzi się, przynajmniej w klasycznych ujęciach Josta Triera i Leo Weisgerbera oraz ich naśladowców, z akceptacji filozoficznych tez Wilhelma Humboldta, dla którego język był „mapą rzeczywistości”, „światem pośredniczącym”, poprzez który użytkownicy języka percyp-

<sup>108</sup> Wybrane zagadnienia teorii pól w językach naturalnych były przedstawione w rozdziale drugim.

<sup>109</sup> Encyklopedia 1993a, s. 406.

<sup>110</sup> Przeciwstawne ujmowaniu słownictwa jako systemu jest traktowanie go jako inwentarza, czyli nieuporządkowanego zbioru form językowych.

pują świat pozajęzykowy. Drugim źródłem koncepcji pola semantycznego były poglądy Ferdynanda de Saussure'a, zwłaszcza jego teoria wartości i sposób rozumienia układu, w którym każda jednostka jest zdeterminowana przez wszystkie pozostałe, a zmiana jednostkowa pociąga za sobą zmianę całego systemu.

Praktyka badań pól częściej koncentrowała się na szczegółowych analizach wybranych fragmentów, a nie na całych systemach leksykalnych<sup>111</sup>. System leksykalny rozpada się na podsystemy – pola znaczeniowe (semantyczne) – w których znaki językowe i ich konfiguracje wzajemnie wyznaczają swoje granice znaczeniowe. Każda zmiana w obrębie pola, np. usunięcie znaku czy wprowadzenie nowego spowoduje semantyczne przekształcenia pozostałych elementów. Kształt poszczególnych pól znaczeniowych, tj. ich zasób leksykalny i wewnętrzne relacje między jednostkami, jest wyrazem właściwego dla danej grupy (lub wszystkich) użytkowników języka rozumienia i ujmowania świata.

W literaturze z zakresu informacji naukowej przyjęła się nieco inna konwencja używania terminów pole semantyczne i system leksykalny. O ile w językoznawstwie mówi się o systemach rozpadających się na podsystemy – pola znaczeniowe (semantyczne), to w nauce o informacji przeważa tendencja do sytuowania tych pojęć na jednym, najwyższym poziomie hierarchii. Mamy zatem system semantyczny lub pole semantyczne jiw, system leksykalny lub pole leksykalne jiw. Jednostki powstałe z podziału systemu (pola) semantycznego najczęściej są nazywane kategoriami lub klasami, przy czym ta ostatnia nazwa, przynajmniej teoretycznie, zarezerwowana jest dla języków charakteryzujących się klasyfikacyjnym podejściem do organizacji słownictwa.

Problemem pól semantycznych w językach informacyjno-wyszukiwawczych zajmował się między innymi Ungurian, rozważając zagadnienia kategoryzacji polskiego nazewnictwa ogólnotechnicznego<sup>112</sup>. Ungurian używał terminu pole semantyczne jiw, definiując go w pewnej opozycji do nieokreślonego pola języka naturalnego. Według Bielickiej i Ścibora polem semantycznym języka informacyjno-wyszukiwawczego jest „wszystko, co można wyrazić za pomocą różnych środków, którymi dany język dysponuje”<sup>113</sup>. Zauważmy, że taka interpretacja *pola semantycznego* wyprowadza nas poza system słownictwa, gdyż pewne znaczenia tworzy się i wyraża wprawdzie na materiale słownikowym, ale poprzez utworzone na mocy reguł gramatycznych znaki wyższego (wyższych) poziomu. Wskazywana przez autorów różnica między systemami języka naturalnego i jiw dotyczy jedynie sposobów ich rozwoju – spontanicznego i niekontrolowanego w przypadku języków naturalnych oraz kontrolowanego i świadomie organizowanego w wypadku jiw.

---

<sup>111</sup> Chcąc być w pełni precyzyjnym, należałoby mówić o podsystemie leksykalnym, pozostawiając cechę 'bycia systemem' samemu językowi.

<sup>112</sup> Ungurian 1982.

<sup>113</sup> Bielicka 1981, s. 30.



Podobnie jak w przypadku języków naturalnych istotnym problemem są granice pola znaczeniowego jiw. Granice pola semantycznego języka informacyjno-wyszukiwawczego wyznacza nie tylko zakres tematyczny dziedziny, dla której język jest tworzony, rozszerzony o niezbędne elementy dziedzin pokrewnych, ale także te cechy, które w rozdziale pierwszym za Blairem zostały nazwane kontekstowymi. Przy określaniu zakresu pola semantycznego nie można ignorować zakresu tematycznego dokumentów (pierwotnych), które są lub będą podstawą tworzenia dokumentów pochodnych, składających się na zbiór informacyjno-wyszukiwawczy danego systemu oraz treści pytań (jeśli są znane), na jakie system ma udzielać odpowiedzi. Zakres tematyczny zbioru dokumentów pierwotnych gromadzonych w systemie zawsze jest jednak szerszy od przyjmowanego zakresu pola semantycznego jiw, ponieważ w jego skład wchodzi elementy pewnych dokumentów nie związane z profilem systemu informacyjno-wyszukiwawczego i, co za tym idzie, podlegające redukcji w procesie indeksowania.

Pole semantyczne języka informacyjno-wyszukiwawczego najlepiej jest analizować, gdy plan wyrażania języka jeszcze nie obowiązuje. Podczas budowy poszczególnych pól dokonuje się nie tylko uzupełnienia i korekty słownictwa, ale i określenia wartości semantycznej poszczególnych jednostek. Należy jednak pamiętać o tym, że uporządkowane w pola semantyczne jednostki leksykalne jiw nawet jeśli są wyrażeniami języka naturalnego, to są nimi jedynie w planie wyrażania, ich plan treści jest bowiem zdefiniowany nie przez reguły semantyczne języka naturalnego, a przez projektanta jiw. Są to więc jednostki quasi-naturalne (paranaturalne) mające formę wyrażen języka naturalnego i znaczenie związane, ale niekoniecznie identyczne, ze znaczeniem odpowiadających im wyrażen języka naturalnego. Zarówno system leksykalny jiw, jak i poszczególne pola znaczeniowe są realnie istniejącymi ugrupowaniami znaków językowych, powstającymi na drodze zespalania elementów znaczeń lub odwrotnie – dzielenia substancji znaczeniowej jiw „od góry”. Wybór metody budowy języka (dedukcyjna czy indukcyjna) nie przesądza o organizacji systemu leksykalnego.

Wydaje się, że systemy leksykalne języków informacyjno-wyszukiwawczych wykazują więcej cech wspólnych z postulatami sformułowanymi przez Triera (ściśle przyleganie do siebie pól, jednoznaczność, określona wartość znaczeniowa i kompleksowy charakter zmian semantycznych) niż systemy języków naturalnych, z myślą o których Trier je sformułował. System Triera był teoretycznym „wzorem” struktury leksykalnej, któremu nie zawsze odpowiadały i do którego nie zawsze dawały się nagiąć zjawiska żywego języka, ale który może stanowić pewien model dla języków informacyjno-wyszukiwawczych. O ile bowiem w wypadku języków naturalnych ma się do czynienia z ogromnym zróżnicowaniem pól semantycznych w idiolektach poszczególnych użytkowników danego języka, o tyle na gruncie tradycyjnych jiw, w tradycyjnych systemach informacyjno-wyszukiwawczych, zjawiska te trak-

tuje się jako marginalne, wychodząc z założenia, że jiw są językami sztucznymi o projektowanej semantyce. Sztuczność semantyki jiw nie powinna jednak oznaczać arbitralności. System leksykalno-semantyczny jiw nie musi i nie powinien być apriorycznym systemem znaczeń.

System leksykalny jiw jako przedmiot projektowania wymaga też opisu sfery interpretacji semantycznej oraz wyboru metody określania znaczeń wyrażen jiw. Omówiona w rozdziale pierwszym Wittgensteinowska koncepcja znaczenia jako użycia może i powinna mieć zastosowanie w projektowaniu jiw jako dopełniająca tradycyjnie stosowane podejście denotacyjne i konotacyjne. Chcąc przedstawić charakterystykę pola semantycznego języka informacyjno-wyszukiwawczego, nie wolno ignorować jego związków z relewantnym fragmentem systemu leksykalno-semantycznego języka naturalnego. Wprawdzie Babik pisze, że „Budowa systemu leksykalnego jiw jest zabiegiem sztucznym lub zabiegiem naturalnym. Gdy projektant danego jiw sam według własnego uznania lub przyjętych a priori zasad tworzy pole semantyczne i buduje system leksykalny danego jiw, wówczas jest to zabieg sztuczny”<sup>114</sup>, ale trudno przypuścić, żeby nawet przy założonej „sztuczności zabiegu” nie odwoływał się on do struktur semantycznych istniejących w pewnym języku naturalnym.

Można zatem powiedzieć, że praca nad systemem leksykalnym języka informacyjno-wyszukiwawczego jest działaniem operującym na dwóch modelach<sup>115</sup>: modelu odwzorowującym odpowiedni fragment systemu leksykalnego języka naturalnego oraz na projektującym modelu systemu leksykalnego jiw. Dobór materiału leksykalnego, wybór koncepcji znaczenia oraz metody analizy znaczeń determinują efekty w postaci określonego modelu systemu leksykalnego jiw. Trudnym problemem jest wybór optymalnej metody analizy semantycznej. Wykorzystanie metody składnikowej wydaje się praktycznie niewykonalne. W wielu sytuacjach użyteczny może być aparat definicyjny oparty na konotacyjnej interpretacji znaczenia i wykorzystujący stosowane w jiw, zwłaszcza w kodach semantycznych, metody opracowania leksyko-graficznego terminów. Funkcję metajęzykową może tu także pełnić język naturalny. Komplementarnie lub równolegle może być użyty język predykatów pierwszego stopnia, umożliwiającą przedstawienie cech danego wyrażenia w postaci zbioru formuł analitycznych. Etap szczegółowego definiowania wyrażen języka naturalnego, choć podstawowy dla budowy systemu leksykalnego jiw, jest często pomijany i zastępowany działaniami związanymi raczej z ustalaniem relacji między jednostkami leksykalnymi jiw. Praktyczną

---

<sup>114</sup> Babik 1996, s. 34.

<sup>115</sup> Za Pszczołowskim (Pszczołowski 1978, s. 119) przyjmijmy, że model jest to „przedmiot złożony (także abstrakcyjny) odwzorowujący dla celów poznawczych lub praktycznych bardziej od niego złożony istniejący albo projektowany fragment rzeczywistości. Pomiędzy dwoma przedmiotami złożonymi (A oraz B), z których jeden jest przedmiotem modelowanym, drugi jego modelem, zachodzi stosunek izomorfizmu struktury (homomorfizmu, izofunkcjonalizmu) tego rodzaju, że po przekształceniu A w przedmiot A' oraz B w przedmiot B' relacja izomorfizmu struktury zostaje zachowana”.

tego konsekwencją może być niedostateczna ostrość wyrażen jiw, trudności z organizacją systemową leksyki i posługiwaniem się językiem w indeksowaniu i wyszukiwaniu.

Ważnym problemem jest wewnętrzny porządek w polach. W ustalaniu porządku pionowego w językach informacyjno-wyszukiwawczych wykorzystuje się albo relacje hierarchiczne, obejmujące, poza zwykłą hipero- i hiponią, także inne charakterystyczne dla semantyki jiw związki o własności przechodniości i przeciwsymetryczności, tj. związek mereologiczny i hierarchii tematycznej, albo, w przypadku jiw, których cały zasób leksykalny (lub jego podzbiory) tworzy szereg<sup>116</sup>, cechy planu wyrażania języka, czyli kolejność znaków w bazie notacyjnej. W określaniu porządku wyrażen w szeregu najczęściej wykorzystuje się również kolejność znaków w bazie notacyjnej. Możliwe są jednak i inne metody. Ungurian scharakteryzował sformułowane przez Ranganathana pragmatyczne, choć niedoskonałe, zasady użytecznego następstwa (ang. *helpful sequence*), wyróżniając wśród nich pięć grup odwzorowujących cechy desygnatów, takie jak:

- czas (późniejszy w czasie, późniejszy w rozwoju ewolucyjnym, itd.);
- lokalizacja przestrzenna (od dołu do góry, od lewej do prawej lub odwrotnie, zgodnie z ruchem wskazówek zegara, od peryferii do środka lub odwrotnie, itd.);
- miara jednostek (zwiększająca lub zmniejszająca się, np. punkt, linia, płaszczyzna, bryła, przestrzeń czterowymiarowa; biblioteka wojewódzka, biblioteka powiatowa, biblioteka miejska, biblioteka dzielnicowa, itd.);
- kompleksowość (np. głoska, sylaba, słowo, itd.);
- inne, obejmujące trzy zasady: następstwa kanonicznego, liczebności publikacji i kolejności alfabetycznej.

Wielu autorów publikacji z zakresu projektowania jiw pisze o kwalifikowaniu gromadzonego słownictwa do „wcześniej zaprojektowanych kategorii semantycznych”<sup>117</sup>, zdając się zakładać, że każdy nowo tworzony język powinien mieć budowę kategoriałną. Skłonni jesteśmy przyznać im rację, mając jednak świadomość, że w praktyce funkcjonuje wiele (może nawet więcej) języków o innej organizacji systemu leksykalnego, którymi w tej pracy jednak się nie zajmujemy. Trudno natomiast znaleźć jakiegokolwiek przemyślenia dotyczące dochodzenia w różnych sytuacjach projektowych do owych kategorii semantycznych. Będzie o tym mowa dalej.

---

<sup>116</sup> Zagadnienia wyboru zasad podziału klasyfikacyjnego, założeń klasyfikacyjnej systematyzacji łańcuchów i szeregów itp. wchodzi w zakres teorii klasyfikacji i tylko pośrednio łączą się z przedmiotem pracy. Dlatego nie będziemy ich omawiać, zwłaszcza że problematyka ta ma liczną literaturę. M.in.: Buchanan 1979; Dobrowolski 1956; Foskett 1963; Loosc 1993; Sosińska 1993; Szrejder 1975; Ungurian 1976. Vickery 1975.

<sup>117</sup> „Gromadzone nazwy z języka naturalnego często są przydzielane do jednej z wcześniej zaprojektowanych kategorii semantycznych, stanowiących szkielet struktury pola semantycznego i szkielet klasyfikacyjny danego jiw”. Babik 1996, s. 42.

W pewnym ujęciu kategoryzacja może być postrzegana jako przetwarzanie bodźców fizycznych, psychiczne, intelektualnych itp. do postaci pojęć (znaczeń). Według Fosketa<sup>118</sup> można wyróżnić pięć następujących kategorii pojęć:

- obiekty (ang. *entities*)
- czynności (ang. *activities*)
- abstrakcje (ang. *abstracts*)
- własności (ang. *properties*)
- kategoria heterogeniczna.

Kategoria obiekty reprezentuje przedmioty fizyczne lub „rzeczy, które możemy wskazać”. Kategoria czynności zawiera (wyrażane za pomocą czasowników i rzeczowników odsłownych) pojęcia identyfikujące czynności, działania i akcje. Kategoria abstrakcje reprezentuje „jakości lub stany”, podczas gdy pojęcia tworzące kategorię własności (wyrażane najczęściej za pomocą połączenia rzeczownika z przymiotnikiem) reprezentują „znaczenia konotacyjne” lub „własności mechaniczne”. Kategoria heterogeniczna składa się z pojęć, w których można wyróżnić składniki semantyczne odpowiadające czterem wcześniejszym kategoriom, ale mimo to postrzegane jako pojęcia samoistne i odrębne. Willetts<sup>119</sup> wyróżnił między innymi następujące pojęcia heterogeniczne:

- Rola człowieka (obiekt+czynność; obiekt+własność), np.: nauczyciel;
- Grupy ludzi (obiekt+abstrakcja), np.: społeczeństwo, uczestnicy konferencji;
- Typ budynku (obiekt+czynność+własność), np.: biblioteka, teatr;
- Dyscypliny (obiekt+czynność+własność+abstrakcja), np.: fizyka, medycyna.

W przypadku języków informacyjno-wyszukiwawczych wewnętrzna struktura kategorii często pozostaje pod wpływem tendencji do systematyzacji, która z kolei odwołuje się do logiki indukcji i dedukcji. W procesie indukcji punktem wyjścia jest identyfikacja pojęcia, następnie można określić ich podstawowe klasy i fasety oraz zgrupować je w klasy czy kategorie wyjściowe (główne). Podejście dedukcyjne implikuje podział zakresu pola semantycznego jiw na z góry założone podpola. Jak wcześniej wspomniano, w literaturze z zakresu języków informacyjno-wyszukiwawczych dość powszechnie można spotkać stwierdzenie, że na etapie opracowywania leksyki jiw zgromadzone słownictwo jest przydzielane do wcześniej zaprojektowanych kategorii semantycznych. Niemal nikt nie pisze jednak, jak owe kategorie semantyczne powinny być lub zostały wyodrębnione, jakie problemy stanęły przed projektantem języka, których z nich nie udało się rozwiązać i dlaczego oraz jakie są tego następstwa. Na przykład stwierdza się jedynie, że „potrzeby związane z zapisem informacji faktograficznej zostały wzięte pod uwagę przy opra-

---

<sup>118</sup> Foskett 1996, s. 77-79.

<sup>119</sup> Willetts 1975, s. 158-184.

cowywaniu kategoryzacji, tj. przy wyborze kategorii semantycznych (kategorii podstawowych i podkategorii)”<sup>120</sup>. Bielicka zaproponowała orientacyjny schemat kategorii, które „nie mają w pełni charakteru uniwersalnego, powinny być dostosowane do określonej dziedziny (wiedzy) lub gałęzi gospodarki narodowej (tematyki tezaurusu)”<sup>121</sup>, wyróżniając:

- dziedziny wiedzy i/lub gałęzie gospodarki narodowej grupujące nazwy „dyscyplin podstawowych i pomocniczych oraz gałęzi gospodarki narodowej”;
- przedmioty, czyli obiekty, obejmujące „nazwy wszystkich rzeczy, czyli wszystko to, co posiada jakieś cechy indywidualne, np. nazwy zwierząt, roślin, związków chemicznych, surowców, produktów, maszyn, części maszyn, narzędzi, przedsiębiorstw, instytucji naukowych, placówek kulturalnych, układów międzynarodowych, pojęć abstrakcyjnych itp. Jeśli jest to celowe, kategorię tę można dzielić na: maszyny, części maszyn, wyroby, instytucje itp.”;
- osoby – „obejmuje nazwiska osób, nazwy zawodów itp. Poza tym do kategorii tej zalicza się również nazwy klas i grup społecznych (np. robotnicy, inteligencja), narodowości i grup etnicznych (Polacy, Mulaci), grup zawodowych (hutnicy) oraz grup charakteryzujących się pewną wspólną sytuacją (inwalidzi, bezrobotni), płcią i/lub wiekiem (dziewczeta, starcy)”;
- procesy, zjawiska, wydarzenia, relacje – „grupuje nazwy wszystkich czynności, procesów, technologii, metod, zjawisk, wydarzeń, relacji itp. Kategorię tę można dzielić dalej, np. można wydzielić relacje, procesy agensywne (procesy, w których występuje sprawca) i procesy bezagensywne (procesy przebiegające samoczynnie) lub procesy fizyczne, chemiczne, biologiczne itp.”;
- własności i wielkości obejmujące „wszystkie deskryptory będące nazwami cech różnych obiektów, a także jednostki miar, wag itp.”;
- prawa – „grupuje nazwy wszystkich praw naukowych, teorii i hipotez dotyczących danej dziedziny, a także przepisy i normy prawne”;
- miejsce obejmuje „nazwy geograficzne i ogólne określenia miejsca (jeziora, góry itp.) potrzebne do określenia miejsca występowania procesu lub obiektu omówionego w dokumencie”;
- czas – „grupuje określenia czasu potrzebne do opisu treści dokumentów, np. daty zawarcia układu, wystąpienia zjawiska, wydarzenia itp., a także ogólne określenia czasu (pory roku, dni, doby i jej części itp.)”.

Generalnie można się spodziewać, że kategorie są ustanawiane na mocy albo apriorycznego, odwołującego się do wiedzy i intuicji projektanta, podziału zakresu pola semantycznego języka, albo analizy denotacji i konotacji zgromadzonych wyrażań. Mało prawdopodobne wydaje się odwoływanie do struktury rzeczywistości dokumentacyjnej, na użytek której język jest tworzony. Wyodrębnianie kategorii nabiera właściwego sensu wtedy, gdy grupując

---

<sup>120</sup> Dao 1989, s. 125.

<sup>121</sup> Bielicka 1988, s. 53-55.

wyrażenia w jednej kategorii, projektant języka informacyjno-wyszukiwawczego bierze pod uwagę nie tylko znaczenia językowe znaków, ale i te relewantne składniki semantyczne, które pojawiają się przy oglądzie rzeczywistości. Dla pewnych systemów i języków informacyjno-wyszukiwawczych *kot* będzie to nie tylko „zwierzę domowe z rodziny *Felis domestica*”, ale także, a może przede wszystkim, „ulubieniec – ulubione zwierzę domowe”<sup>122</sup>; *czosnek* to nie tylko „roślina warzywna cebulkowa z rodziny liliowatych”, ale także „przyprawa kuchenna” i/lub „naturalny środek leczniczy” i/lub „coś, co ma przykry zapach”.

W rozdziale drugim scharakteryzowaliśmy utworzoną przez Rosch koncepcję podstawowego poziomu kategoryzacji. Podstawowe poziomy kategoryzacji to struktury uznane przez użytkowników języka za najbardziej użyteczne. Stanowią one punkt wyjścia do tworzenia dalszych kategorii – poziomów nadrzędnych i podrzędnych. Według Rosch w leksykonie każdego języka naturalnego status poziomu podstawowego uzyskuje to piętro hierarchii, które okazuje się najużyteczniejsze w codziennej komunikacji, najskuteczniejsze i najbardziej ekonomiczne w użyciu. Na podstawie poglądów Rosch na naturę kategorii prototypowych i postulowanego przez nią istotnego kognitywnie i językowo podstawowego poziomu kategoryzacji można sformułować tezę, że słownik języka informacyjno-wyszukiwawczego, dążąc do osiągnięcia właściwego poziomu szczegółowości jednostek leksykalnych, nie może się obejść bez pojęć przynależących do poziomu podstawowego. Należy jednak zwrócić uwagę na fakt, że kategorie podstawowe (poziomu podstawowego), chociaż najbliższe największej liczbie ludzi, nie są uniwersalne. Jest prawdopodobne, że dla ornitologa pojęciem z poziomu podstawowego kategorii będzie ŚWISTUNKA PIECUSZEK, a dla dziecka PTAK. Podstawowy – lub nie – charakter kategorii zależy od doświadczenia osobniczego, także od doświadczeń kulturowych, społecznych, wykształcenia itp. użytkowników języka informacyjno-wyszukiwawczego. Szczególnie wyraziście można to zaobserwować w przypadku pojęć abstrakcyjnych. Kategorie podstawowe mają także, a może przede wszystkim, zastosowanie w wyszukiwaniu. Użytkownicy formułując pytania wykazują skłonność do posługiwania się pojęciami poziomu podstawowego nawet wtedy, gdy faktyczny obiekt ich zainteresowania lokuje się na innym poziomie. Posługiwanie się pojęciami poziomu podstawowego może jednak prowadzić do wyszukania dużych ilości informacji, których selekcja wymagałaby stosowania odpowiednich strategii wyszukiwawczych sprzężonych z regułami tworzenia reprezentacji dokumentów i utrzymywania zbiorów informacyjnych systemu.

Określenie podstawowego dla projektowanego języka informacyjno-wyszukiwawczego poziomu kategoryzacji powinno pozytywnie wpływać na, na ogół intuicyjnie rozumiany, poziom szczegółowości jednostek leksy-

---

<sup>122</sup> Zauważmy, że kategoria ULUBIENIEC według Wierzbickiej nie byłaby kategorią taksonomiczną, a funkcjonalną, jak ZABAWKA czy BRON.

kalnych. W rozprawie nie będziemy się zajmować problemem szczegółowości, który tylko pośrednio łączy się z kategoryzacją. Właściwa szczegółowość słownictwa jiw jest jednym z ważniejszych, choć ciągle niedocenianych i nie zbadanych aspektów języka informacyjno-wyszukiwawczego. Zwykle rozumie się, jakie mogą być konsekwencje zbyt małej szczegółowości. Ale bardzo rzadko projektanci jiw zastanawiają się nad tym, jakie negatywne skutki może powodować szczegółowość nadmierna (por. s. 161), zwłaszcza jeśli w słowniku danego jiw albo nie określa się znaczeń poszczególnych jednostek, albo określenia te są z różnych powodów „niedoskonałe”. Na przykład w słowniku języka haseł przedmiotowych Biblioteki Kongresu między innymi można znaleźć cztery następujące tematy: **Husband and wife, Marriage, Married people, Marital status**<sup>123</sup>. Jedyne przy Husband and wife znajduje się pewna bezpośrednia wskazówka co do jego znaczenia i stosowania. Mówi ona, że temat Husband and wife stosuje się do prac na temat prawnych aspektów związków łączących męża i żonę. Znaczenia pozostałych haseł można jedynie próbować „zrekonstruować” na podstawie ekwiwalentów wyszukiwawczych, miejsca w systemie paradygmatycznym języka oraz użyć tych terminów w charakterystykach wyszukiwawczych dokumentów, na przykład w katalogu Biblioteki Kongresu. Mimo to trudno byłoby wskazać ostre zakresy znaczenia i stosowania haseł. Wątpliwe wydaje się, żeby skutecznie mógł to zrobić użytkownik, którego wiedza o istocie języka informacyjno-wyszukiwawczego i regułach posługiwania się nim w tworzeniu reprezentacji dokumentów i w wyszukiwaniu zwykle jest bardzo skromna. Ponadto użytkownik niewiele może wiedzieć o przedmiocie swoich poszukiwań. Dla bibliotekarzy i projektantów jiw powinien to być dodatkowo ważny sygnał, że dokumenty o podobnej treści mogą być (i są) reprezentowane przez różne hasła przedmiotowe.

Analizując zagadnienie tworzenia kategorii w językach informacyjno-wyszukiwawczych, chcemy zwrócić uwagę na dotąd praktycznie niewykorzystywaną w tym celu, wypracowaną przez językoznawstwo kwantytatywne, metodę ciągów definicyjnych. W metodzie tej wykorzystuje się hierarchiczne własności pojęć odkryte w wyniku badania prawidłowości statystycznych jednostek języka i tekstu. Wywodzące się od Arystotelesa podejście klasyczne wyrażające się we wskazaniu cech istotnych (koniecznych i wystarczających), wspólnych dla wszystkich desygnatów wyrażenia, charakteryzuje się następującymi własnościami: wyrażenie definiowane reprezentuje całą klasę przedmiotów i stanowi abstrakcję cech istotnych tej klasy; cecha istotna jest cechą konieczną, a zbiór wszystkich cech istotnych stanowi warunek wystarczający do zdefiniowania; terminy podrzędne charakteryzują się wszystkimi cechami istotnymi (definicyjnymi) terminu nadrzędnego i dodatkowo cechami istotnymi dla właściwego im poziomu taksonomicznego, co pozwala na konstruowanie drzew hierarchicznych terminów. Ciągi definicyjne są to ciągi hiperonimów wyodrębnianych z kolejnych definicji leksemu (traktowanego

---

<sup>123</sup> LCSH 1998 , T.2. s. 2659-2660, T. 3. s. 3470, 3476-3477, 3479.

jako leksem wyjściowy) poprzez poszukiwanie członów oznaczających genus proximum. Analizy ciągów definicyjnych słowników różnych języków naturalnych, w tym języka polskiego, zmierzały do wykrycia uniwersalnych składników semantycznych, mogących stanowić podstawowe elementy meta-języka opisu znaczenia. Są nimi końcowe człony ciągów, zwykle nie mające w danym słowniku definicji ze względu na swą ogólność, mogące wyznaczać generalny podział kategoryalny analizowanego uniwersum leksyki języka.

Metodą ciągów można badać zarówno leksemy jedno- jak i wieloznaczne. Te ostatnie prowadzą do powstania gniazd leksykalnych, czyli struktur, które buduje się na podstawie ciągów definicyjnych uwzględniających wszystkie znaczenia leksemu w badanych definicjach. Do tej pory metodę ciągów definicyjnych stosuje się niemal wyłącznie do analiz ogólnych słowników językowych. Badania nad możliwością wykorzystania metody ciągów definicyjnych do słownictwa specjalistycznego w systemach informacyjno-wyszukiwawczych prowadzili między innymi: Elżbieta Piechocka<sup>124</sup>, Dorota Głazek<sup>125</sup> i Wiesław Babik<sup>126</sup>. Badania te, choć o różnych celach i stopniu trudności, pozwoliły przede wszystkim na ocenę jakości badanych słowników terminologicznych oraz (badania Babika) możliwości wykorzystania metody ciągów definicyjnych do komputerowego wspomaganie projektowania jiw. Metodę tę, a właściwie opracowany na jej podstawie metajęzyk opisu znaczeń, można by stosować również jako element warsztatu leksykograficznego przy tworzeniu słowników terminologicznych. Z punktu widzenia zagadnienia kategoryzacji w jiw najciekawsze są analizy przeprowadzone przez Babika, który dokonał analizy jakości (spójności systemów definicji, odesłań i terminologicznego) *Słownika pięcioletniego z zakresu fotogrametrii i teledetekcji* oraz zbadał możliwości jego wykorzystania do automatycznego wygenerowania języka informacyjno-wyszukiwawczego. Szukał odpowiedzi m.in. na pytania: Jaką strukturę ma pole semantyczne badanego słownika? Jakie są możliwe do wygenerowania typy struktur pola semantycznego jiw i jakich typów języków informacyjno-wyszukiwawczych? Jaka jest wartość wygenerowanych struktur? W rezultacie stwierdził, że metoda ciągów definicyjnych „umożliwia wyjście w budowie systemu leksykalnego jiw od najniższego poziomu, to jest poziomu terminów, na podstawie którego można by stopniowo budować strukturę pola semantycznego jiw, przechodząc od terminów hasłowych (niekoniecznie charakteryzujących się najniższym stopniem ogólności, lecz w każdym razie stopniem ogólności niższym od najwyższego), do wyższych stopni ogólności”<sup>127</sup>. W przypadku *Słownika pięcioletniego z zakresu fotogrametrii i teledetekcji* na najwyższym poziomie ogólności (siódmym) znalazł się jeden termin – GRUPA ZNAKÓW<sup>128</sup>, na poziomie szóstym nie

---

<sup>124</sup> Picchocka 1992.

<sup>125</sup> Bojar 1998.

<sup>126</sup> Babik 1996.

<sup>127</sup> Ibidem s. 153.

<sup>128</sup> Zachowujemy tu oryginalną formę typograficzną.



wyodrębniono żadnej jednostki leksykalnej, na poziomie piątym – CZĘŚĆ WYPOSAŻENIA, ZJAWISKO, LOKALIZACJA, KORPUS, INFORMACJA, na czwartym – URZĄDZENIE, NAZWA OBRAZU, CZĘŚĆ PRZESTRZENI, NAZWA, TERMIN, PODZBIÓR OBRAZU, ZABURZENIE, CZĄSTECZKA, OTRZYMYWANIE itd. „Terminy na najwyższych poziomach ogólności nadają się na nazwy kategorii semantycznych. Pozostałe elementy ciągów, a szczególnie terminy znajdujące się w bezpośrednim sąsiedztwie terminów końcowych, mogą zostać wykorzystane do budowy kolejnych stopni (uszczegółowiania) pola semantycznego jiw. Wspólny element ciągów definicyjnych terminów posiadających na danym poziomie te same terminy może stać się nazwą podkategorii. Taka procedura może być kontynuowana aż do momentu, gdy w wyróżnionych w ten sposób kategoriach n-tego stopnia podziału uzyskamy pojedyncze drzewka hierarchiczne kończące się terminami hasłowymi (wyjściowymi)”<sup>129</sup>. Przytoczone przykłady wprowadzicie nie w pełni potwierdzają zacytowany wniosek, nie zmniejsza to jednak potencjalnej użyteczności metody. Automatyczne generowanie całego zasobu słownictwa jiw i automatyczne ustanawianie systemu kategorii i siatki relacji paradygmatycznych metodą ciągów definicyjnych jeszcze nie jest chyba możliwe. Metodę tę można byłoby jednak potraktować jako komputerowe wspomaganie niektórych etapów projektowania języka informacyjno-wyszukiwawczego łączącego „inteligencję ludzką z <<maszynową>> w interaktywnym procesie budowy jiw”<sup>130</sup>. Analizując terminy końcowe (nie zdefiniowane w *Słowniku...*), Babik zredukował otrzymane gniazda terminologiczne do następujących kategorii: CZYNNOŚCI, DZIEDZINY NAUKI, JEDNOSTKI, MATERIAŁ, METODA, OBIEKT (OBRAZ, KĄT, UKŁAD, SYSTEM, PUNKT), OBSZAR, PŁASZCZYZNA, POWIERZCHNIA, PROCES, PRZESTRZEŃ, PRZYRZĄDY, RELACJA (WIELKOŚĆ, WŁAŚCIWOŚĆ, STOSUNEK, ZALEŻNOŚĆ), URZĄDZENIA. „Wyrażenia te mogłyby pełnić funkcję nazw kategorii semantycznych grupujących terminy o wspólnym elemencie znaczeniowym, tworząc jednocześnie odpowiednie kategorie leksykalne jiw”<sup>131</sup>. Należy przy tym pamiętać, że jakość otrzymanej struktury w największym stopniu jest determinowana kompletnością słownika będącego źródłem materiału leksykalnego oraz typem i jakością zastosowanych w nim definicji. Obecnie nie jest możliwe wykorzystanie metody ciągów definicyjnych do automatycznego generowania języków informacyjno-wyszukiwawczych. Można jednak ją stosować do wspomagania procesu projektowania, o ile w danej dziedzinie istnieją dobre słowniki, zawierające relewantną informację.

Pewne, dające się rozszerzyć na języki informacyjno-wyszukiwawcze, refleksje dotyczące budowania systemu kategorii można znaleźć w literaturze

---

<sup>129</sup> Babik 1996, s. 108.

<sup>130</sup> Bojar 1998, s. 20.

<sup>131</sup> Babik 1996, s. 122.

z zakresu sztucznej inteligencji i inżynierii wiedzy<sup>132</sup>. Daje się tu zauważyć analogia z problemami projektowania języków i systemów informacyjno-wyszukiwawczych. Jako przykład przytoczymy, zachowując specyficzną terminologię, zalecenia dotyczące budowy tzw. ontologii wyrażone przez Uscholda i Gruningera<sup>133</sup>. Przez ontologię w sztucznej inteligencji i inżynierii wiedzy najczęściej rozumie się schemat konceptualny systemu, dostarczający logicznego opisu danych, umożliwiający współdziałanie programów aplikacyjnych i baz danych bez konieczności zapewniania wspólnych struktur danych na poziomie symbolicznym. Ontologia dostarcza słownictwa pozwalającego mówić o dziedzinie, podczas gdy odpowiednia baza symbolicznie reprezentuje wiedzę konieczną do rozwiązywania problemów.

Według Uscholda i Gruningera najpierw należy wybrać dziedzinę, która ma być objęta ontologią. Terminy gromadzone są poprzez badanie literatury z wybranej dziedziny oraz konsultowanie się ze specjalistami. Uschold i Gruninger zalecają zastosowanie metody „burzy mózgów” z udziałem ekspertów w celu ustalenia znaczących terminów oraz ich względnej wagi. Zgromadzone terminy należy następnie tak pogrupować, by były one bardziej powiązane z terminami w ramach grupy niż z terminami z innych grup. W trakcie tego procesu można odnotować potencjalne synonimy oraz odesłania semantyczne, czyli pojęcia, do których prawdopodobnie wystąpią odwołania lub które będą przywoływane przez pojęcia w innych grupach.

Kolejny etap to formułowanie definicji. W związku z tym Uschold i Gruninger wprowadzili pojęcie metaontologii, przez które rozumieją system kategorii służących do wzajemnego wiązania terminów. W systemie ENTERPRISE ontologia składa się z JEDNOSTEK (ang. *Entities*) oraz zbioru RELACJI (ang. *Relations*) między JEDNOSTKAMI. Jednostki pełnią określone ROLE (ang. *Roles*) w ramach RELACJI. Szczególnym rodzajem RELACJI jest ATRYBUT (ang. *Attribute*), zaś szczególnym rodzajem ROLI jest DZIAŁAJĄCY (ang. *Actor*). STAN RZECZY (ang. *State-of-affairs*) jest sytuacją, w której dowolna kombinacja JEDNOSTEK tworzy dowolną liczbę ich wzajemnych RELACJI. Na przykład:

JEDNOSTKA: *PLAN*

RELACJA: *POSIADAĆ ZDOLNOŚĆ* jest RELACJĄ między *OSOBA* i *DZIAŁALNOŚCIĄ*;

*SPRZEDAŻ* jest RELACJĄ między *JEDNOSTKAMI PRAWNYMI* służącą wymianie *PRODUKTU* za *CENĘ* sprzedaży;

ROLA: *SPRZEDAWCA* jest ROLĄ odgrywaną przez *JEDNOSTKĘ PRAWNĄ* w RELACJI *SPRZEDAŻ*;

ATRYBUT: *DATA URODZENIA* jest ATRYBUTEM łączącym tylko jedną *DATE* z daną *OSOBA*.

<sup>132</sup> Gruber 1995; Guarino 1995; Sowa 1995; Uschold 1996.

<sup>133</sup> Uschold 1996.

Uschold i Gruninger zalecają, żeby początkowo nie wiązać się z żadną konkretną metaontologią<sup>134</sup>. Najpierw należy dokładnie rozważyć znaczenia i ich wzajemne powiązania, aby w efekcie sformułować wymagania wobec metaontologii. Z kolei tworząc definicje, należy rozpocząć od tych obszarów, które w największym stopniu pokrywają się semantycznie z innymi obszarami. Następnie zdefiniować terminy podstawowe<sup>135</sup> w każdym obszarze. Autorzy nazywają to podejście „wypośrodkowanym”, a jako przykład podają *pies*. Jeśli termin ten potraktujemy jako podstawowy, wówczas *ssak* jest generalizacją (uogólnieniem), a *spaniel* – specjalizacją. Komentują to następująco: podejście „od dołu w górę” gwarantuje wysoki stopień szczegółowości, ale utrudnia ustalenie tego co wspólne między powiązаныmi pojęciami. Z kolei podejście „od góry w dół” może doprowadzić do wyboru i arbitralnego narzucenia metaontologicznych kategorii wysokiego poziomu, mogących stać się swoistym ograniczeniem.

Na użytek ontologii systemu ENTERPRISE sformułowali następujące cztery klasy terminów, które mogą być potraktowane, nawiązując do terminologii języków informacyjno-wyszukiwawczych, jako ulegające dalszemu podziałowi kategorie specjalistyczne:

#### DZIAŁANIA I PROCESY

Działalność  
 Specyfikacja-działalności  
 Wykonać  
 Rozpoczęcie-zadania  
 Zakończenie-zadania  
 Warunki wstępne/konieczne  
 Efekt  
 Wykonawca  
 Pod-działalność  
 Władza  
 Właściciel działalności  
 Zdarzenie  
 Plan  
 Subplan  
 Planowanie  
 Specyfikacja-problemu  
 Zdolność  
 Wprawność  
 Zasoby  
 Rozmieszczenie zasobów  
 Zamiana-zasobów

#### ORGANIZACJA

Osoba  
 Maszyna  
 Korporacja  
 Partnerstwo  
 Partner  
 Jednostka prawna  
 Jednostka-organizacyjna  
 Kierowanie  
 Delegowanie/pełnomocnictwo  
 Ogniw-zarządzania  
 Własność-prawna  
 Własność-nieprawna  
 Właściciel  
 Majątek/kapitał  
 Udziałowiec  
 Umowa o zatrudnieniu  
 Akcja  
 Akcjonariusz

<sup>134</sup> W literaturze z zakresu informacji naukowej (np.: Ranganathan 1967; Ungurian 1983, 1982, 1976) przeprowadzono wiele analogicznych dyskusji, nazywając je technikami analizy fascetowej oraz kształtowania kategorii fascetowych.

<sup>135</sup> Terminy podstawowe u Uscholda i Gruningera wykazują wiele cech wspólnych z kategoriami poziomu podstawowego u Rosch.

## STRATEGIA

Cel  
Utrzymanie-celów  
Cel-zamierzony  
Utrzymujący-cel  
Cel-strategiczny  
Cel/przedmiot  
Wizja  
Misja  
Cel/zadanie  
Uzyskanie-pomocy  
Strategia  
Planowanie-strategii  
Działanie-strategiczne  
Decyzja  
Założenie  
Założenie-krytyczne  
Założenie-niekrytyczne  
Czynnik-wpływu  
Krytyczny-czynnik-sukcesu  
Ryzyko  
Promocja  
Konkurent

## MARKETING

Sprzedaż  
Potencjał-sprzedaży  
Na-sprzedaż  
Oferta-sprzedaży  
Sprzedający  
Aktualny-klient  
Potencjalny-klient  
Detalista  
Produkt  
Żądanie-ceny  
Cena-sprzedaży  
Rynek  
Segment-ryнку  
Zmienna-segmentacji  
Badania-ryнку  
Marka towaru  
Wizerunek  
Charakterystyka  
Zapotrzebowanie  
Potrzeby-ryнку

Brian Vickery<sup>136</sup> przytacza za Maheshem ważniejsze zasady „inżynierii ontologii”. Zauważmy, że przynajmniej na takim poziomie werbalizacji wykazują one duże podobieństwo do wymagań stawianych systemom kategorii w jiw. Są to:

– nieunikatowość: nie istnieje unikatowa (niepowtarzalna) ontologia. Ontologie nie są tworam naturalnymi, które wystarczy odkryć, lecz artefaktami do skonstruowania, których struktura będzie zależeć od zamierzonego użycia;

– zależność zadaniowa: ontologię buduje się do konkretnych celów. Ontologia budowana do przetwarzania języka naturalnego może być nieodpowiednia dla innych celów, takich jak planowanie, projektowanie, czy inny rodzaj wnioskowania. Rodzaje wiedzy, które musi zawierać dana ontologia, są różne dla poszczególnych zadań<sup>137</sup>;

<sup>136</sup> Vickery 1997.

<sup>137</sup> Nie ma jednak zgody co do tego warunku. Jedni autorzy podkreślają wagę zapewnienia możliwie dużej niezależności ontologii od zadań (dla celów ponownego wykorzystania), co oznacza, że możliwe jest dojście do jednolitej ontologii pewnego podzbioru wiedzy. Inni z kolei wskazują na konieczność podporządkowania ontologii zadaniom, twierdząc, że nie istnieje jednolita ontologia. Istnieje zatem konflikt między dążeniem do generalizacji i specjalizacją. Znaczny nacisk na potencjalne wielokrotne wykorzystanie ontologii, a także ich wartość jako ogniwa wiążącego w komunikacji między systemami, przemawiają za tym, by ontologie były niezależne od zadań. Jednakże, podobnie jak i w naszej dziedzinie, chęć ukształtowania ontologii zgodnic

– łatwość wykorzystania: ontologia powinna być łatwa do przeglądania, znajdowania w niej pojęć, przedstawiania związków danego pojęcia z innymi pojęciami;

– modularność: musi istnieć możliwość narastającego dodawania nowych pojęć i relacji;

– ziarnistość (rozdrobnienie): budowanie precyzyjnej ontologii, w której każde pojęcie jest wyraźnie odróżnialne od innych pojęć, wymaga „drobnoziarnistej” dekompozycji pojęć. Wybór poziomu rozdrobnienia musi być określony na podstawie użyteczności dla postawionego zadania;

– redundancja: klasyfikacja<sup>138</sup> pojęć w ontologii jest z konieczności redundantna. Różne jej wymiary często się nakładają. Nie należy dążyć do całkowitego wyeliminowania redundancji.

Analogia z językami i systemami informacyjno-wyszukiwawczymi jest tu oczywista, choć nie można nie widzieć też różnic, wynikających prawdopodobnie z różnych celów ontologii i jiw. Inżynierowie wiedzy nie sięgają do dorobku teorii i praktyki języków i systemów informacyjno-wyszukiwawczych lub czynią to bardzo rzadko (i vice versa). Oczywiście jest również powiązanie ontologii z leksykografią, zwłaszcza że kilka ontologii zostało zaprojektowanych do wykorzystania w przekładzie maszynowym. Jednak w publikacjach z zakresu inżynierii wiedzy rzadko jawnie nawiązuje się do dorobku leksykologii i językoznawstwa. Nie widać też zbyt dużego zrozumienia faktu, że budowa ontologii nie jest ani nowym rodzajem działalności, ani refleksji teoretycznej. Należy mieć nadzieję, że rozwój narzędzi inżynierii wiedzy spowoduje większe zrozumienie wagi analizy semantycznej w przetwarzaniu informacji oraz stworzy nowe możliwości rozwiązywania problemów reprezentacji i wyszukiwania informacji.

## KATEGORYZOWANIE INFORMACJI

Kategoryzacja jest również pojmowana jako proces przyporządkowywania pewnych obiektów kategoriom już istniejącym w języku informacyjno-wyszukiwawczym. Bywa wówczas określana mianem *kategoryzowania*. Aby uczynić powyższe stwierdzenie ostrym, należy dodać, że nie chodzi tu

---

z potrzebami określonej grupy użytkowników przemawia za tym, że rozwiązanie to powinno być skrojone „na miarę”. Problem ten znany jest od dawna w teorii klasyfikacji. Twórcy klasyfikacji „starają się skonstruować standardowy schemat, gdy można założyć wspólnotę celów wykorzystania, lecz przedkładają zróżnicowane schematy, by sprostac zróżnicowanym celom. Wiąże się to z (...) potrzebą włączenia subschematów zawierających te same terminy, lecz umieszczone w różnych kontekstach”. Vickery 1997, s. 285.

<sup>138</sup> Terminy redundancja i klasyfikacja zostały użyte przez Vickery’ego w znaczeniach innych niż te, które są im najczęściej przypisywane. Przez klasyfikację Vickery rozumie tu dość dowolny podział zbioru znaczeń lub wyrażen języka; redundancja zaś oznacza wielokrotne występowanie pewnych znaczeń lub wyrażen w różnych klasach.

o wprowadzanie nowych znaczeń i wyrażeń do struktury jiw, ale o kategoryzację dokumentów, a dokładniej informacji w nich zawartych i informacji o dokumentach. Jeśli przyjąć, że dokumentem jest utrwalona informacja<sup>139</sup>, to pytania użytkowników (choć utrwalone nie zawsze w obiektywnie postrzeganej formie), będące podstawą formułowania instrukcji wyszukiwawczych określających dystynktywne cechy informacji relewantne dla danego użytkownika, mogą być traktowane jako podklasa dokumentów, co pozwala na rozszerzenie poczynionych obserwacji i wniosków także na pytania użytkowników.

Obraz świata zawarty w dokumencie jest wyrazem, między innymi, przyjętej przez autora dokumentu kategoryzacji tego świata. Język informacyjno-wyszukiwawczy jest z kolei obrazem porządku kategoryjnego przyjętego przez jego projektantów. Użycie jiw polega na jak najlepszym „wpisaniu” obrazu świata dokumentu w obraz rzeczywistości ukształtowany w jiw. W tym sensie kategoryzacja przestaje być operacją wewnątrzjęzykową, stając się zarówno działaniem, jak i jego rezultatem łączącym dokument i język informacyjno-wyszukiwawczy. Tak rozumiana *kategoryzacja* wykazuje pokrewieństwo semantyczne z *indeksowaniem*. Chcąc bliżej określić ten związek, zaczniemy od uprawdopodobnienia tezy, że *kategoryzowanie* i *indeksowanie* nie są terminami równoznacznymi. Opowiadając się za ich niesynonimicznością, odwołamy się jedynie do istoty ustroju pola semantycznego pewnych języków informacyjno-wyszukiwawczych. Fałszywie brzmiałoby nazywanie kategoryzowaniem tworzenia reprezentacji dokumentu w języku o niekategoryjnej strukturze pola semantycznego, na przykład w języku słów kluczowych. Wprawdzie i wówczas można mówić o kategoryzacji, ale ma ona charakter niejawny i nierелеwantny z punktu widzenia przedmiotu i celu rozprawy, będąc rezultatem istnienia i oddziaływania osobniczego kategoryjnego obrazu świata na płaską strukturę języka informacyjno-wyszukiwawczego. Z drugiej strony tradycja językowa sprawia, że używanie języków informacyjno-wyszukiwawczych typu klasyfikacji fasetowej, najbliższe znaczeniu *kategoryzowania*, nazywane jest klasyfikacją lub klasyfikowaniem. Nie można jednak wykluczyć, że zrozumienie istoty kategoryzacji i istoty języków informacyjno-wyszukiwawczych o kategoryjnej strukturze systemu leksykalno-semantycznego może wpłynąć na zmianę normy terminologicznej.

Skoro *kategoryzowanie* i *indeksowanie* pozostają w niesynonimicznym związku znaczeniowym, to ważna staje się odpowiedź na pytanie, czy między zakresami tych terminów zachodzi związek inkluzji czy krzyżowania. W ujęciu modelowym indeksowanie jest procesem, na który składają się następujące po sobie czynności analizowania, selekcyjonowania i tłumaczenia. Analizowane są dokumenty, zwłaszcza zawarte w nich informacje. Ponieważ nie wszystkie

---

<sup>139</sup> Dokument – utrwalona informacja (wraz z materiałem, w którym została utrwalona, nośnikiem informacji). Słownik 1993a, s. 35.

informacje są ważne z punktu widzenia kryteriów relewancji danego systemu informacyjno-wyszukiwawczego, dlatego pewne z nich są eliminowane z dalszego przetwarzania. Jest to czynność selekcjonowania informacji. W efekcie tych dwóch czynności powstaje pewien obraz treści dokumentu i jego samego. Obraz ten może przyjąć jawną formę tekstu języka naturalnego, choć nie zawsze tego samego, co język dokumentu. Tekst ten jest interpretacją informacji zawartej w dokumencie i wyrazem postrzegania dokumentu przez osobę indeksującą. On też może być przedmiotem pierwszego tłumaczenia na język swobodnych słów kluczowych, stając się podstawą utworzenia charakterystyki słownej dokumentu (instrukcji słownej pytania). Charakterystyka słowna tłumaczona jest z kolei na obowiązujący w systemie język informacyjno-wyszukiwawczy, prowadząc do powstania charakterystyki wyszukiwawczej dokumentu (instrukcji wyszukiwawczej pytania) zwanej też reprezentacją dokumentu (pytania). W praktyce granice między poszczególnymi etapami nie są ostre, a sekwencja czynności ulega wielokrotnym powtórzeniom. Dodatkowo na rezultaty prowadzonych działań oddziałują liczne, trudne do zidentyfikowania i słabo poddające się badaniu, czynniki o charakterze indywidualnym, osobniczym, w tym idiosynkratyczne systemy kategorii osób tworzących reprezentacje dokumentów. Czy zatem cały ten proces, angażujący także gramatykę i pragmatykę jiw, można nazwać kategoryzacją, czy tylko jakąś jego część? A jeśli tak, to którą? Naszym zdaniem, miano kategoryzacji przysługuje całemu procesowi tworzenia reprezentacji, ponieważ w wielu językach informacyjno-wyszukiwawczych kategoryzacja leksykalna i syntaktyczna są nierozzerwalnie ze sobą połączone, wzajemnie się uzupełniając i oddziałując jedna na drugą. Należy jednak pamiętać, że:

- używany język informacyjno-wyszukiwawczy powinien być „językiem z kategoriami”;
- obowiązująca obecnie norma językowa, uwzględniająca inne aspekty rzeczywistości, sprawia, że kategoryzacja bywa zwykle określana innymi mianami.

Podobieństwo kategoryzacji i indeksowania sprawia, że wiele, często do dziś nie rozwiązanych problemów indeksowania odnosi się również do kategoryzacji. Wskażemy niektóre z nich, gdyż współtworzą one kontekst, w którym przebiega kategoryzacja i powstają reprezentacje dokumentów i pytań. Większość tych problemów wymaga odrębnych badań, bez których trudno sobie wyobrazić dalszy rozwój dyscypliny i praktycznej działalności informacyjnej.

Jakość reprezentacji dokumentów w systemie informacyjno-wyszukiwawczym ma niekwestionowany wpływ na efektywność systemu. Ważnym warunkiem jakości odwzorowania informacji relewantnej jest istnienie w języku odpowiednich (relewantnych) i właściwie ustrukturalizowanych klas semantycznych „wypełnionych” jednostkami leksykalnymi, co do których istnieje przekonanie, że są lub będą potrzebne do charakteryzowania dokumen-

tów. Kategorialne podejście do organizacji pola semantycznego jiw daje projektantowi języka możliwość wskazania tych cech semantycznych wyrażen jiw, które przy podejściu denotacyjno-konotacyjnym byłyby niemożliwe do uwzględnienia. Każdej klasie elementów rzeczywistości wyodrębnionej na podstawie pewnego zespołu cech odpowiada odrębne znaczenie w systemie znaczeń. Ten sam wycinek rzeczywistości może być zatem prezentowany za pomocą różnych znaczeń i ich układów, ale każdemu znaczeniu powinien odpowiadać tylko jeden znak językowy. Wskazywanie w znaczeniu wyrażenia jiw na podstawie różnych kryteriów relewancji różnych cech i atrybutów obiektów rzeczywistości dokumentacyjnej i pozadokumentacyjnej sprawia, że obiekty te mogą być różnie kategoryzowane, tj. kwalifikowane do różnych kategorii, zależnie od tego, która z cech jest w danym przypadku cechą relewantną, na przykład **Tytoń** może się znaleźć zarówno w kategorii **Rośliny psiankowate**, jak i **Używki**, **Czynniki kancerogenne**, **Surowce przemysłowe**; termin **Pomidor** może być postrzegany zarówno jako element kategorii **Rośliny psiankowate**, jak i kategorii **Rośliny warzywne**, **Rośliny uprawne**, **Surowiec przemysłowy**, **Surowiec do przetworów domowych** itd.

Brak relewantnych kategorii, także brak jednostek leksykalnych w istniejących kategoriach, nie mogą nie mieć wpływu na jakość reprezentacji dokumentów. Warunek istnienia w języku informacyjno-wyszukiwawczym kategorii relewantnych należy zatem postrzegać co najmniej w trzech aspektach – zupełności systemu kategorii i wypełniających je wyrażen jiw, szczególności kategorii i wyrażen jiw oraz umiejscowienia kategorii względem siebie w systemie leksykalno-semantycznym języka i poszczególnych wyrażen (znaczeń) w konkretnych kategoriach. Zagadnienia te powinny stać się przedmiotem wnikliwych i kompleksowych badań, w rezultacie których zostałyby sformułowane (na ile możliwe) mocne przesłanki do projektowania systemów kategorii w językach informacyjno-wyszukiwawczych. Warto zauważyć, że zwłaszcza trzeci z aspektów może być źródłem problemów z „dopasowywaniem” obrazu rzeczywistości zawartego w dokumencie i w języku informacyjno-wyszukiwawczym. Na przykład, zajęcie Wilna przez armię sowiecką w 1939 roku w języku haseł przedmiotowych Biblioteki Kongresu zostało zakwalifikowane jako okupacja rosyjska, podczas gdy w jhp Instytutu Informacji Naukowej Nauk Społecznych Akademii Nauk ZSRR (INION) – jako Wielka wojna ojczyźniana narodu radzieckiego.

Kolejny warunek jakości reprezentacji dokumentów ma charakter pozajęzykowy i dotyczy dwóch aspektów tworzenia charakterystyk wyszukiwawczych – szerokości (zwanej też kompletnością) i szczególności (głębokości). Postulat niepominiania informacji relewantnych nie jest równoznaczny z koniecznością wprowadzania do charakterystyki dokumentu wszystkich jego wątków tematycznych w stopniu proporcjonalnym do tego, jaki przedstawia dokument. Pominiecie informacji relewantnej prowadzi wprawdzie do strat informacji, ale zbyt szerokie reprezentacje, uwzględniające informacje



o znaczeniu marginalnym dla danej grupy użytkowników, mogą powodować wzrost szumu informacyjnego. Poziom szerokości, podobnie jak poziom szczegółowości, powinien być ustalany w efekcie badania systemu i oczekiwań użytkowników. Wprawdzie ma on charakter na ogół indywidualny, ale i tu istnieje potrzeba i możliwość badań ogólnych, na przykład wyznaczających metodologię prowadzenia badań indywidualnych.

Analizując strukturę procesu tworzenia reprezentacji dokumentów nie można nie zauważyć potencjalnych źródeł błędów. Występują one na każdym etapie i często nie są dostrzegane przez osoby tworzące reprezentacje. Błędnie przeprowadzona analiza dokumentu, nadmierna lub zbyt mała selekcja informacji, brak syntezy informacji to tylko niektóre z problemów. W aspekcie kategoryzacji zagrożeniem dla poprawności reprezentacji jest (obok braku relevantnych kategorii) nieprzystawalność czy częściowa przystawalność obrazów świata dokumentu i języka informacyjno-wyszukiwawczego. O ile autor dokumentu na ogół dąży do indywidualnego kształtowania obrazu interesującej go rzeczywistości, o tyle jiw zmierza w kierunku pewnej obiektywizacji tego obrazu. Naukowcy, zajmując się tym samym wycinkiem rzeczywistości, tworzą różne, niekiedy sprzeczne, jego interpretacje i obrazy. Projektanci jiw z kolei tworząc język informacyjno-wyszukiwawczy albo szukają rozwiązania kompromisowego, albo wybierają jedną z wielu możliwych koncepcji, teorii itp., przejmując także określony porządek kategoryalny. Nieprzystawalność obrazów świata odwzorowanych w różnych jiw (podobnie jak dzieje się to w różnych językach naturalnych) jest też przyczyną kłopotów z przekładalnością (kompatybilnością) jiw.

Istnieje wreszcie związek między kategoryzacją jako rodzajem ustrukturalizowanej organizacji pola semantycznego jiw a sprawnością wyszukiwawczą systemu. Pomijając te aspekty, które wiążą się z jakością reprezentacji, nie można nie zwrócić uwagi na to, że w językach informacyjno-wyszukiwawczych o strukturze płaskiej interpretacja cech semantycznych wyrażen jiw opiera się na założeniu, że znaczenia te są identyczne ze znaczeniami równokształtnych wyrażen języka naturalnego i że użytkownicy komunikując się z systemem, będą potrafili przewidzieć sposoby wyrażania interesujących ich znaczeń przez autorów poszukiwanych dokumentów i autorów charakterystyk tych dokumentów. Jak pokazują badania, zróżnicowanie form prezentacji tych samych znaczeń w tekstach dokumentów jest bardzo duże, a umiejętność ich przewidzenia nie jest powszechna. Na przykład z badań Saltona<sup>140</sup> wynika, że wprawdzie 75% dokumentów wyszukanych na podstawie słów kluczowych zostało uznane za relevantne, ale tylko 20% posiadanych przez system dokumentów relevantnych zostało wyszukanych.

Podsumowując, należy stwierdzić, że istnieje potrzeba i konieczność dalszych badań kategoryzacji, które w pierwszej kolejności powinny dotyczyć dwóch aspektów tego zagadnienia: jakości odwzorowania informacji relevant-

---

<sup>140</sup> Salton 1986.

nej w charakterystykach dokumentów (jakości charakterystyk) oraz jakości wyszukiwania w związku z kategoryzacją w jiw.

Tymczasem w praktyce biblioteczno-informacyjnej *kategoryzacja*, rozumiana jako termin pokrewny *indeksowaniu*, zwykle oznacza kategoryzację dokumentów zawierających teksty literackie. Najczęstsze kryteria, według których literatura piękna jest kategoryzowana w systemach informacyjno-wyszukiwawczych, są następujące:

- przynależność etniczna, narodowa lub państwowa autora (tzw. aspekt osoby);
- język, w którym został napisany, względnie z którego lub na który został przetłumaczony, dany utwór literacki (tzw. aspekt języka);
- forma literacka (rodzaj, gatunek, odmiana gatunkowa, tzw. aspekt formy piśmienniczej);
- okresy literackie lub historyczne (tzw. aspekt czasu);
- kraj, w którym rozwija się dana literatura (tzw. aspekt miejsca).

Mamy zatem literaturę polską, literaturę amerykańską, literaturę szwajcarską, literaturę amerykańską w języku hiszpańskim, literaturę żydowską w Polsce międzywojennej, literaturę żydowską w języku polskim, literaturę szwajcarską w języku niemieckim, literaturę średniowieczną, poezję nowożytną, powieść wojenną, nowele, opowiadania, miniatury literackie itp. Praktyczne funkcjonowanie wyodrębnionych pięciu kategorii informacji o tekstach literackich pozostaje w ścisłym związku z naturą kategoryzowanej rzeczywistości i oczekiwań użytkowników konkretnego systemu informacyjno-wyszukiwawczego. Wbrew pozorom kategoryzacja tekstów literackich nie jest czynnością mechaniczną. Problemy, jakie się mogą wówczas pojawić, dotyczą wielu grup zagadnień, spośród których ważniejszymi wydają się być: forma tekstu literackiego, przynależność do okresu literackiego lub historycznego, przynależność autora do grupy językowej, etnicznej, państwowej oraz kompletność i szczegółowość kategoryzacji.

W zasadzie większych problemów z kategoryzacją nie przysparza jedynie konwencjonalna powieść dziewiętnastowieczna. Natomiast współczesne teksty literackie, choć oczywiście nie wszystkie, a także teksty pochodzące z epok, kiedy funkcjonowały inne typy genologiczne, charakteryzują się (z różną siłą) cechami wskazującymi na związek z wieloma kategoriami, jak ma to miejsce w przypadku kategorii kognitywnych, w tym kategorii prototypowych. Trudność sprawia jednoznaczne przypisanie wielu tekstów literackich do jednego określonego gatunku lub tzw. odmiany gatunkowej. Niejednokrotnie trudno też jest orzec, czy dany tekst jest powieścią, opowiadaniem, minipowieścią itp., gdyż ani terminologia fachowa, ani sam tekst nie dostarczają wyrazistych (dystynktywnych) kryteriów porównania i klasyfikacji. Na przykład powieść *Trzej muszkieterowie* Aleksandra Dumasa najczęściej jest kategoryzowana jako historyczna i/lub awanturnicza i/lub sensacyjna. Podobne problemy pojawią się przy kategoryzowaniu literatury w aspekcie osoby (autora) ze względu na dwujęzyczność pewnych autorów, brak określenia przynależności do narodu czy

grupy etnicznej lub, przeciwnie, eksponowanie owej przynależności pozostającej w pewnej sprzeczności z innymi cechami, na przykład z używanym językiem. Złożonym problemem jest też kategoryzacja w aspekcie czasu. Wydaje się, że warunki klasycznej kategoryzacji mogą spełniać jedynie kategorie języka i miejsca.

Ważnym zagadnieniem jest szerokość i szczegółowość kategoryzacji tekstów literackich. Szerokość jest, ogólnie biorąc, pytaniem o to, jakie kategorie informacji o tekście literackim powinny być odwzorowane w jego charakterystyce. Niżej wymieniamy najczęstsze kryteria kategoryzacji literatury pięknej w systemach informacyjnych. Nie znaczy to, że zawsze tam, gdzie się ma do czynienia z kategoryzacją tekstów literackich wszystkie te kryteria są czy powinny być stosowane. Uznawanie pewnych z nich za obowiązkowe, innych zaś za fakultatywne czy zbędne należy do sfery pragmatyki języka informacyjno-wyszukiwawczego. Bardzo odczuwa się tu brak badań przydatności wyszukiwawczej różnych kryteriów kategoryzacji literatury.

Jedną z ciekawszych propozycji opracowania literatury pięknej jest koncepcja Walkera<sup>141</sup>. Walker zastosował opracowaną przez Ranganthana formułę PMEST (por. s. 132) do opisu dokumentu zawierającego tekst literacki, następująco interpretując poszczególne kategorie:

- P [Personality] – katalogowanie autorsko-tytułowe,
- M [Matter] – przedmiot tekstu literackiego, tematyka utworu,
- E [Energy] – typ tekstu ze względu na formę piśmienniczą, przynależność gatunkową itp.,
- ST [Space Time] – autor według języka i okresu.

Następnie w kategorii M wyodrębnił, także według schematu PMEST, podkategorie, a wynik tej przeprowadzonej na dwóch poziomach kategoryzacji wyraził za pomocą rozszerzonej formuły PM(PMEST)E(ST), której elementy uzyskały następującą interpretację:

\* M(PMEST) – przedmiot tekstu literackiego w tym:

- P – postaci występujące w utworze,
- M – środowisko społeczne, psychiczne, zawodowe, kulturowe itp. (społeczność, rodzina, małżeństwo, dzieci, więźniowie, kategorie osób, także etniczne, społeczne, zawodowe, warunki życia, zwyczaje, obyczaje, konflikty, relacje z... itd.),
- E – temat, idea przewodnia utworu,
- S – umiejscowienie geograficzne,
- T – umiejscowienie czasowe.

Pytanie, czy i w jakim stopniu należałoby się stosować do zaleceń Walkera oraz ewentualnie jak określać wartości poszczególnych kategorii i podkategorii, jest problemem, który musi być rozwiązany odrębnie w każdym systemie informacyjno-wyszukiwawczym. Ze znanych bibliotek formułą Walkera stosuje lub co najmniej nawiązuje do niej Biblioteka Kongresu.

---

<sup>141</sup> Walker 1958.

Ważnym problemem kategoryzacji tekstów literackich jest jej szczegółowość. Spróbujemy nieco bliżej scharakteryzować to zagadnienie na przykładzie kategorii Forma literacka. Z jednej strony problem ten pozostaje pod wpływem warunków systemu informacyjno-wyszukiwawczego (na przykład, czy z punktu widzenia użytkowników ważne jest, żeby *Zamek kaniowski* Seweryna Goszczyńskiego był identyfikowany jako „romantyczna powieść poetycka”, czy wystarczy informacja „powieść poetycka”, a może „powieść romantyczna”), z drugiej zaś szczegółowości słownictwa danego języka informacyjno-wyszukiwawczego. Brak w języku wyrażen o odpowiedniej szczegółowości, a co za tym idzie, sile selekcyjnej, uniemożliwia adekwatną reprezentację dokumentów. Z drugiej strony nadmiar wyrażen, w tym wyrażen szczegółowych, powoduje problemy, o których była już mowa na s. 160 i 181-182. Na podstawie *Słownika literatury popularnej*<sup>142</sup>, przez analogię i zgodnie z zasadą, że tworząc język informacyjno-wyszukiwawczy, bierze się pod uwagę możliwość istnienia desygnatów wyrażen, a nie ich faktyczne istnienie, można utworzyć (wprawdzie nie według jednego kryterium) nazwy następujących typów powieści: autobiograficzna, awanturyczna, biograficzna, dydaktyczna, dla dzieci, dla dziewcząt (pensjonarska), dla kobiet, dla młodzieży, dziennikarska, dżungli, erotyczna, fantastyczna, fantastycznonaukowa, fantasy, feministyczna, filozoficzna, gangsterska, grozy, historyczna, horror, humorystyczna, kryminalna, marynistyczna, miłosna, moralizatorska, obyczajowa, podróżnicza, polityczna, przygodowa, przyrodnicza, psychologiczna, religijna, rodzinna, romans, romans zbójcecki, rycerska, satyryczna, sensacyjna, społeczna, szpiegowska, tajemnic, thriller, western, wojenna, zbójcka. Tworząc język informacyjno-wyszukiwawczy i organizujące go kategorie semantyczno-leksykalne, należałoby ocenić, czy wszystkie wyróżnione tu odmiany gatunkowe powieści będą użyteczne zarówno z punktu widzenia potrzeb użytkowników, jak i własności zgromadzonych dokumentów. Należy przy tym pamiętać:

- o możliwych trudnościach ze wskazaniem cech dystynktywnych, na mocy których granice kategorii miałyby ostry charakter; kategorie typów powieści (choć nie tylko one wśród tekstów literackich) zdają się mieć naturę bliższą kategoriom prototypowym niż Arystotelesowskiem;
- że same teksty literackie rzadko reprezentują „czyste” typy formy.

Nie będziemy rozwijać wątku kategoryzacji tekstów literackich<sup>143</sup>, gdyż jest to obszerny i złożony temat, który łatwo można zbanalizować, poddając powierzchownej analizie. W przeciwieństwie do powszechnie obserwowanej praktyki, uważamy, że zagadnienie reprezentacji tekstów literackich stanowi ważny, choć niedoceniany, problem. Mamy nadzieję, że stanie się on przedmiotem zarówno wnikliwych analiz praktycznych, jak i pogłębionej refleksji teoretycznej.

<sup>142</sup> Słownik 1997.

<sup>143</sup> O problemach tych pisali między innymi: Baker 1988, Beghtol 1986, Pejtersen 1983; 1979, Spiller 1980. Pewne elementy i aspekty tego problemu zostały też zbadane przez Stanisławę Kurck-Kokocińską (Kurck-Kokocińska 1999).

## WNIOSKI

Celem przeprowadzonych badań nad kategoryzacją w językach informacyjno-wyszukiwawczych było przede wszystkim dobre, wielostronne określenie problemu badawczego traktowane jako pierwszy i podstawowy etap wszelkich innych działań. Termin kategoryzacja używany był w rozprawie w trzech znaczeniach: tworzenie kategorii w języku informacyjno-wyszukiwawczym; rezultat tak rozumianej kategoryzacji, czyli system kategorii; przydzielanie nowych elementów do obowiązującego w jiw systemu kategorii. Obecnie na podstawie przeprowadzonych analiz i obserwacji chciałabym sformułować kilka wniosków ogólnych.

Nadal jednym z najślabszych elementów teorii i praktyki języków informacyjno-wyszukiwawczych jest niewystarczająca wiedza o bezpośredniej rzeczywistości pozajęzykowej, czyli o świecie dokumentów. Wiedza ta powinna być jedną z dwóch fundamentalnych podstaw orzekania o jiw. Często się zapomina, że organizacja słownictwa odwołująca się do elementów znaczeniowych wymaga z jednej strony wskazania najistotniejszych cech, które powinny być odwzorowane w strukturze morfologicznej wyrażenia lub w jego definicji, z drugiej zaś – wskazania cech uznanych za relewantne ze względu na funkcje metainformacyjną i wyszukiwaczą jiw. Projektanci jiw większą wagę przypisują relewantności cech obiektów rzeczywistości pośrednio denotowanej przez jiw, nie doceniając w pełni wagi cech obiektów bezpośrednio wskazywanych przez jiw, czyli dokumentów. Tymczasem nie da się poprawnie i wyczerpująco wskazać relewantnych cech konotacyjnych i denotacyjnych znaków jiw bez odwoływania się między innymi do wiedzy o rzeczywistości dokumentów i funkcjonowania jiw. Znaczenie znaku języka informacyjno-wyszukiwawczego, podobnie jak znaczenie znaku języka naturalnego, to nie (nawet najwierniejsze) odzwierciedlenie cech odpowiadającego wyrażeniu obiektu, ale świadoma, wsparta na regułach relewancji transformacja językowa uwydatniająca jedne cechy, a pomniejszająca, czy wręcz eliminująca inne.

Istotnym, choć niedocenianym, elementem reprezentacji dokumentów są informacje kontekstowe. Informacje o kontekście dotyczące przynajmniej typowych sytuacji problemowych i działań, w wyniku których powstają lub w trakcie których wykorzystywane są określone dokumenty, powinny być uwzględniane przy projektowaniu jiw i tworzeniu reprezentacji dokumentów. Należy przy tym pamiętać, że informacje kontekstowe nie zawsze są jawnie zawarte w dokumentach. Czasami można je zidentyfikować jedynie znając sytuację, w których określone dokumenty znajdują właściwe zastosowanie.

Prawdopodobne wydaje się, że dodatni wpływ na funkcjonowanie języków informacyjno-wyszukiwawczych mogłoby mieć połączenie konotacyjno-denotacyjnej koncepcji znaczenia z Wittgensteinowskim postulatem obserwowania używania znaków językowych jako podstawy do wyrokowania o znaczeniach. Przeszkodą może być spotykane w praktyce oddzielanie projektowania jiw od jego używania, co, zwłaszcza w wypadku języków budowanych metodą dedukcyjną, może sprawiać, że projektant języka nie jest w stanie wskazać przykładów użycia znaków jiw. Przykładami tymi są bowiem dokumenty, w których charakterystykach wyszukiwawczych wystąpił dany znak jiw. To właśnie dokumenty nadają znakom języka rzeczywiste znaczenia, niekiedy różniące się od znaczeń słownikowych, tj. przewidzianych przez projektanta. Innym sposobem wykorzystania koncepcji znaczenia jako użycia mogłoby być udostępnianie informacji o przebiegu wybranych realnych wyszukiwań. Wprawdzie wiele wyszukiwań jest unikatowych i trudno porównywalnych, ale udostępniane przykłady można byłoby traktować jako ilustracje pewnych regularności zachodzących w systemie i w języku informacyjno-wyszukiwawczym – to, jak może wyglądać dobre wyszukiwanie, jak posługiwać się wyrażeniami jiw, jak budować z nich zdania i teksty, jak reagować na różne „zachowania” systemu itp. (ang. *teaching by example*). Tradycyjnie rozwiązuje się te problemy, zapewniając użytkownikowi pomoc osoby pośredniczącej, wykwalifikowanego bibliotekarza czy pracownika informacji, który zna system, jiw, techniki i strategie wyszukiwawcze, często ma nawet głęboką wiedzę o systemie, ale na ogół mniejszą o przedmiocie wyszukiwania, a na pewno nie ma tej świadomości potrzeb i oczekiwań wobec systemu, którą ma użytkownik.

Ze względu na istnienie ścisłego związku między kategoryzacją w jiw a efektywnością systemu informacyjno-wyszukiwawczego istnieje potrzeba i konieczność podjęcia badań kategoryzacji, które, ogólnie biorąc, powinny dotyczyć dwóch aspektów – jakości reprezentacji dokumentów, w tym jakości odwzorowania kategorii przedstawionych w dokumencie i w języku informacyjno-wyszukiwawczym oraz jakości wyszukiwania w związku z kategoryzacją w jiw. Jednym z warunków wysokiej jakości odwzorowania informacji jest istnienie w języku informacyjno-wyszukiwawczym odpowiednich (relevantnych) klas semantycznych zawierających jednostki leksykalne potrzebne do charakteryzowania dokumentów. Kategoriale podejście do organizacji pola

semantycznego jiw daje projektantowi języka możliwość wskazania tych cech semantycznych wyrażen jiw, które przy podejściu denotacyjno-konotacyjnym byłyby trudne lub wręcz niemożliwe do ustalenia. Każdej klasie elementów rzeczywistości wyodrębnionej na podstawie pewnego zespołu cech odpowiada odrębne pojęcie, a zatem ten sam wycinek rzeczywistości może być prezentowany w jiw za pomocą różnych wyrażen i ich układów. Wyodrębnianie w znaczeniu wyrażenia jiw na podstawie różnych kryteriów relewancji różnych cech i atrybutów obiektów rzeczywistości dokumentacyjnej i pozadokumentacyjnej sprawia, że obiekty te mogą być kwalifikowane do różnych kategorii, zależnie od tego, która z cech brana jest pod uwagę. Brak relewantnych kategorii, także brak jednostek leksykalnych w istniejących kategoriach, nie mogą mieć wpływu na jakość reprezentacji dokumentów. Istnienie w języku informacyjno-wyszukiwawczym kategorii należałoby zatem widzieć i badać co najmniej w dwóch aspektach – zupełności systemu kategorii i wypełniających je wyrażen jiw oraz szczegółowości kategorii i wyrażen jiw.

Na efektywność systemu informacyjno-wyszukiwawczego znaczący wpływ mają także szerokość i szczegółowość reprezentacji dokumentów, zwane w literaturze także szerokością i szczegółowością indeksowania. Poziom szerokości, podobnie jak poziom szczegółowości, powinien być ustalany w efekcie badania systemu i oczekiwań użytkowników. Wprawdzie ma on charakter na ogół indywidualny, ale i tu istnieje możliwość i potrzeba badań ogólnych, na przykład wyznaczających metodologię prowadzenia badań indywidualnych.

Istnieje wreszcie związek między kategoryzacją jako rodzajem ustrukturalizowanej organizacji pola semantycznego jiw a sprawnością wyszukiwawczą systemu. Pomijając te aspekty, które wiążą się z jakością reprezentacji, nie można nie zwrócić uwagi na to, że brak kategoryzacji jako formy semantycznej organizacji informacji prowadzi do zmniejszenia się szansy, że użytkownik będzie w stanie identyfikować i interpretować znaczenia znaków jiw oraz efektywnie posługiwać się nimi w procesie wyszukiwania. Wzrost efektywności języków informacyjno-wyszukiwawczych w istotnym stopniu zależy zatem od kategoryzacji jako modelu organizacji systemu leksykalnego jiw i porządku znaczeń znaków językowych.

Autorzy publikacji często stosują termin *kategoryzacja* wymiennie z *klasyfikacja*. Precyzyjne rozdzielenie znaczeń tych dwóch terminów i, co za tym idzie, zakresów ich stosowania nie jest obecnie możliwe, co widać szczególnie wyraźnie w przypadku klasycznego ujęcia kategoryzacji. Klasyfikacja i kategoryzacja mają wiele wspólnych cech i problemów. Klasyfikowanie, odwołując się do rozgałęzionej struktury spełniającej warunki podziału logicznego, implikuje postrzeganie cech niezbędnych i wystarczających do podziału na rozłączne klasy w ramach zupełnej struktury hierarchicznej nałożonej poprzez określone uporządkowanie pewnej rzeczywistości. Z kolei kategoryzacja odnosi się do podziału elementów uniwersum na zespoły przed-

miotów – kategorie – których okazy są wzajemnie powiązane jakąś postrzeżaną, relewantną relacją podobieństwa. W przeciwieństwie do klasyfikacji, kategoryzacja, oczywiście poza ujęciem klasycznym, nie implikuje ani tego, że przynależność do kategorii jest wyznaczona przez postrzeżenie zbioru ustalonych cech, ani tego, by zawieranie się w jednej kategorii wykluczało przynależność do innej. Uznanie podobieństw między skądinąd niepodobnymi jednostkami i wynikająca stąd identyfikacja kategorii pozwalają konstruować różne, odpowiednie do różnych sytuacji, obrazy rzeczywistości i jej uporządkowania.

Należy jednak pamiętać, że system kategorii może skutecznie funkcjonować wówczas, jeżeli nie zmienia się drastycznie za każdym razem, gdy pojawiają się nowe informacje. Równocześnie powinien być wystarczająco elastyczny, aby móc się dostosowywać do zmian otoczenia. Z pewnego punktu widzenia kategoriami najdoskonalszymi, opartymi na całkowitej korelacji cech wszystkich okazów kategorii, są kategorie klasyczne, Arystotelesowskie. Wymaganie całkowitej korelacji cech sprawia jednak, że kategorie klasyczne są dość sztywne i oporne na zmiany. W pewnych sytuacjach kategorie prototypowe mogą skutecznie zastąpić kategorie Arystotelesowskie. Centralne okazy kategorii prototypowych mają dużą liczbę wspólnych atrybutów i pod tym względem centrum kategorii prototypowej zbliża się do ideału kategorii klasycznej. Równocześnie kategorie prototypowe umożliwiają przynależność obiektom, które mają mniej czy niewiele atrybutów wspólnych z okazami centralnymi, ale w świadomości użytkowników języka są z nimi kojarzone.

Wybór organizacji systemu leksykalnego jiw powinien być rezultatem wnikliwej analizy kontekstu, warunków, w jakich będzie stosowany język informacyjno-wyszukiwawczy, a nie decyzją aprioryczną. Projektanci języków informacyjno-wyszukiwawczych powinni też zwracać uwagę na jawne wiązanie modelu kategorialnego projektowanego jiw z zasadnym i wiarygodnym modelem rzeczywistości, którą język będzie opisywał. Nie jest możliwe stworzenie systemu kategorii bez oglądu rzeczywistości. Kategorie, przynajmniej te wykorzystywane w jiw, nie są tworam, które wystarczy odkryć, lecz artefaktami, które należy skonstruować stosownie do zamierzonego użycia.

Na koniec chciałabym przypomnieć, że nie było moim zadaniem stworzenie konkretnej kategoryzacji czy określenie wszystkich możliwych implikacji przeniesienia dorobku różnych dyscyplin naukowych (głównie językoznawstwa) na grunt nauki o informacji. Wysiłki skoncentrowałam na wielostronnym i wnikliwym scharakteryzowaniu dostrzeżonego problemu. Kategoryzacja, stanowiąc o strukturze pola semantycznego języka informacyjno-wyszukiwawczego, jest ważna nie tylko dla teorii jiw, ale również dla praktyki informacyjnej, decyduje bowiem o zawartości zbioru wyszukiwawczego systemu, determinuje postać tworzących go charakterystyk wyszukiwawczych oraz wyznacza stopień przystawalności kategoryzacji informacji w systemie do kate-



goryzacji informacji charakterystycznej dla użytkowników systemu. W pracy podjęłam też próbę uporządkowania i podsumowania wiedzy o językach informacyjno-wyszukiwawczych w zakresie określonym przez przedmiot rozprawy. Można powiedzieć, że praca stanowi wprowadzenie do obszernej problematyki kategoryzacji w jiw, porządkując dotychczasowy stan wiedzy, uzasadniając potrzebę, możliwości i kierunki dalszych badań oraz proponując nowy, trochę odmienny od dotychczasowego, sposób postrzegania rzeczywistości języków i systemów informacyjno-wyszukiwawczych.

Niekiedy dużo dzieli wysiłki poznawcze od praktyki i praktykę od nauki, zwłaszcza w tak szybko i istotnie zmieniającej się rzeczywistości, jak rzeczywistość systemów i języków informacyjno-wyszukiwawczych. Parafrazując myśl Tadeusza Kotarbińskiego<sup>1</sup>, chciałabym zakończyć stwierdzeniem, że, choć czasami nie natychmiastowy i nieoczywisty, istnieje jednak związek między zastosowaniem metod czysto poznawczych a rozwojem praktyki.

---

<sup>1</sup> „(...) rozsądny umiar każe nam zauważyć fundamentalną zależność metod projektowania nowych urządzeń od zastosowania metod czysto poznawczych”. Cyt. za: Sitarska 1990, s. 169.

## BIBLIOGRAFIA ZAŁĄCZNIKOWA

### **Allen 1991a**

Allen Bryce L.: *Cognitive research in information science* // Annual Review of Information Science and Technology Vol. 26 (1991) s. 3-37

### **Allen 1991b**

Allen Bryce L.: *Topic knowledge and online catalog search formulation* // Library Quarterly Vol. 61 (1991) nr 2 s. 188-213

### **Allwood 1990**

Allwood Carl Martin, Wang Zhong-Ming: *Conceptions of computers among students in China and Sweden* // Computers in Human Behaviour Vol. 6 (1990) nr 2 s. 185-199

### **Amerykańska 1993**

*Amerykańska antropologia kognitywna : poznanie, język, klasyfikacja i kultura* / wybór i red. nauk. Michał Buchowski. – Warszawa : Wydawnictwo Instytutu Kultury, 1993

### **Anderson 1986**

Anderson J., Durand Jacques: *Dependency phonology* // Dependency and non-linear phonology / ed. Jacques Durand. – London : Croom Helm, 1986 s. 1-54

### **Apresjan 1980**

Apresjan Jurij D.: *Semantyka leksykalna : synonimiczne środki języka*. – Wrocław : Ossolineum, 1980

### **Artowicz 1997**

Artowicz Elżbieta: *Reprezentacja wiedzy w systemie informacyjno-wyszukiwawczym : zagadnienia relewancji*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1997

### **Artowicz 1990**

Artowicz Elżbieta: *Wiedza o języku oraz o rzeczywistości pozajęzykowej w systemie informacyjno-wyszukiwawczym : rozprawa doktorska wykonana pod kier. doc. dr hab. Bożenny Bojar*. – Warszawa : Uniwersytet Warszawski. Wydział Neofilologii, 1990. – [Maszynopis]

### **Artowicz 1986**

Artowicz Elżbieta, Chmielewska-Gorczyca Ewa: *Wykorzystanie aparatu definicyjnego języków informacyjno-wyszukiwawczych w słownikach terminologicznych* // Zagadnienia Informatyki Naukowej (1986) nr 1 s. 61-82

**Arystoteles 1990a**

Arystoteles: *Kategorie* // Dzieła wszystkie / Arystoteles. T.1. – Warszawa : Państwowe Wydawnictwo Naukowe, 1990 s. 32-63

**Arystoteles 1990b**

Arystoteles: *Topiki* // Dzieła wszystkie / Arystoteles. T.1. – Warszawa : Państwowe Wydawnictwo Naukowe, 1990 s. 343-473

**Babik 1996**

Babik Wiesław: *Generowanie języków informacyjno-wyszukiwawczych ze słowników terminologicznych.* – Kraków : Wydawnictwo Uniwersytetu Jagiellońskiego, 1996

**Babik 1992**

Babik Wiesław: *Struktura pola semantycznego języka informacyjno-wyszukiwawczego dla materiałów kartograficznych w świetle potrzeb ich użytkowników* // *Zagadnienia Informatyki Naukowej* (1992) nr 2 s. 47-65

**Baker 1988**

Baker Sharon L.: *Will fiction classification schemes increase use?* // *Reference Quarterly* Vol. 27 (1988) nr 3 s. 366-376

**Barnett 1988**

Barnett George A., Siegel Gary: *The diffusion of computer-assisted legal research systems* // *Journal of the American Society for Information Science* Vol. 39 (1988) nr 4 s. 224-234

**Bartol-Jarosińska 1996**

Bartol-Jarosińska Danuta: *Zagadnienia leksykalno-semantyczne* // *Nauka o języku dla polonistów / pod red. Stanisława Dubisza.* – Warszawa : Książka i Wiedza, 1996 s. 305-330

**Bates 1979a**

Bates Marcia J.: *Idea tactics* // *Journal of the American Society for Information Science* Vol. 30 (1979) nr 5 s. 280-289

**Bates 1979b**

Bates Marcia J.: *Information search tactics* // *Journal of the American Society for Information Science* Vol. 30 (1979) nr 4 s. 205-214

**Beghtol 1986**

Beghtol Clare: *Bibliographic classification theory and text linguistics : aboutness analysis, intertextuality and the cognitive act of classifying documents* // *Journal of Documentation* Vol. 42 (1986) nr 2 s. 84-113

**Belkin 1990**

Belkin Nicholas J.: *Cognitive viewpoint in information retrieval* // *Journal of Information Science* Vol. 16 (1990) nr 1 s. 11-16

**Belkin 1988**

Belkin Nicholas J.: *On the nature and function of explanation in intelligent information retrieval* // *Proceedings of the 11th International Conference on Research and Development in Information Retrieval*, 1988 June 13-15

Grenoble, France / ed. Y. Chiaramella. – Grenoble : Presses Universitaires de Grenoble, 1988 s. 135-195

**Belkin 1987**

Belkin Nicholas J.: *Discourse analysis of human information interaction for specification of human-computer information interaction* // Canadian Journal of Information Science Vol. 12 (1987) nr 1 s. 31-42

**Belkin 1986**

Belkin Nicholas J., Kwasnik Barbara H.: *Using structural representations of anomalous states of knowledge for choosing document retrieval strategies* // Association for Computing Machinery (ACM) Conference on Research and Development in Information Retrieval, 1986 September 8-10 Pisa, Italy / ed. F. Rabitti. – Baltimore : Association for Computing Machinery, 1986 s. 11-22

**Belkin 1984**

Belkin Nicholas J.: *Cognitive models and information transfer* // Social Science Information Studies Vol. 4 (1984) nr 2/3 s. 112-129

**Belkin 1983**

Belkin Nicholas J., Seeger Thomas, Wersig Gernot: *Distributed expert problem treatment as a model for information system analysis and design* // Journal of Information Science Vol. 5 (1983) nr 4 s. 153-168

**Belkin 1982**

Belkin Nicholas J., Oddy Robert N., Brooks Helen M.: *ASK for information retrieval*. P. 1: *Background and theory* ; P. 2: *Results of a design study* // Journal of Documentation Vol. 38 (1982) nr 2 s. 61-71 ; nr 3 s. 145-164

**Bielicka 1988**

Bielicka Lucyna A.: *Metodyka i organizacja opracowania tezauryusa*. – Warszawa : Centrum Informacji Naukowej, Technicznej i Ekonomicznej, 1988

**Bielicka 1984**

Bielicka Lucyna: *Języki informacyjno-wyszukiwawcze typu obiekt-atrybut systemów faktograficznych*. – Warszawa : Instytut Informacji Naukowej, Technicznej i Ekonomicznej, 1984

**Bielicka 1981**

Bielicka Lucyna, Ścibor Eugeniusz: *Wprowadzenie do teorii języków informacyjnych*. – Warszawa : Centrum Informacji Naukowej, Technicznej i Ekonomicznej, 1981

**Blair 1990**

Blair David C.: *Language and representation in information retrieval*. – Amsterdam : Elsevier, 1990

**Bojar 1998**

Bojar Bożenna, Babik Wiesław: *Metoda ciągów definicyjnych – możliwości jej wykorzystania w systemach informacyjno-wyszukiwawczych* // Zagadnienia Informacji Naukowej (1998) nr 1 s. 3-22

**Bojar 1979**

Bojar Bożenna: *Opis semantyczny czasowników ruchu oraz pojęć związanych z ruchem.* – Warszawa : Wydawnictwa Uniwersytetu Warszawskiego, 1979

**Bojar 1976**

Bojar Bożenna: *O metainformacji i metajęzyku // Zagadnienia Informatyki Naukowej* (1976) nr 2 s. 49-58

**Bookstein 1975**

Bookstein Abraham, Swanson R. Don: *A decision theoretic foundation for indexing // Journal of the American Society for Information Science* Vol. 26 (1975) nr 1 s. 45-50

**Borgman 1986a**

Borgman Christine L.: *Human-computer interaction with information retrieval systems : understanding complex communication behavior // Progress in Communication Sciences* Vol. 7 (1986) nr 2 s. 91-122

**Borgman 1986b**

Borgman Christine L.: *The user's mental model of an information retrieval system : an experiment on a prototype online catalog // International Journal of Man-Machine Studies* Vol. 24 (1986) nr 1 s. 47-64

**Borgman 1984**

Borgman Christine L.: *Psychological research in human-computer interaction // Annual Review of Information Science and Technology* Vol. 19 (1984) s. 33-64

**Brooks 1986**

Brooks Helen M., Daniels Penny J., Belkin Nicholas J.: *Research on information interaction and intelligent information provision mechanisms // Journal of Information Science* Vol. 12 (1986) nr 1/2 s. 37-41

**Buchanan 1979**

Buchanan B.: *Theory of library classification.* – London : Clive Bingley, 1979

**Bunting 1989**

Bunting Karl-Dieter: *Wstęp do lingwistyki.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1989

**Buttler 1967**

Buttler Danuta: *Koncepcje pola znaczeniowego // Przegląd Humanistyczny* R. 11 (1967) z. 2 s. 41-59

**Canter 1985**

Canter David, Rivers Rod, Storrs Graham: *Characterizing user navigation through complex data structures // Behaviour and Information Technology* Vol. 4 (1985) nr 2 s. 93-102

**Caplan 1996**

Caplan David: *Language : structure, processing, and disorders.* – Cambridge : Massachusetts Institute of Technology Press, 1996

**Case 1991**

Case Donald Owen: *Conceptual organization and retrieval of text by historians : the role of memory and metaphor* // Journal of the American Society for Information Science Vol. 42 (1991) nr 9 s. 657-668

**Chmielewska-Gorczyca 1981**

Chmielewska-Gorczyca Ewa: *Relacje syntagmatyczne w językach informacyjno-wyszukiwawczych* : rozprawa doktorska wykonana pod kier. doc. dr hab. Bożenny Bojar. – Warszawa : Uniwersytet Warszawski. Wydział Neofilologii, 1981. – [Maszynopis]

**Chomsky 1982**

Chomsky Noam: *Zagadnienia teorii składni*. – Wrocław : Ossolineum, 1982

**Cleverdon 1970**

Cleverdon Cyril W.: *Progress in documentation : evaluation tests of information retrieval systems* // Journal of Documentation Vol. 26 (1970) nr 1 s. 55-67

**Cleverdon 1967**

Cleverdon Cyril W.: *The Cranfield tests on index language devices* // ASLIB Proceedings Vol. 19 (1967) nr 6 s. 173-193

**Cleverdon 1964**

Cleverdon Cyril W.: *Evaluation of operational information retrieval systems*. P. 1: *Identification of criteria*. – Cranfield, 1964

**Cognition 1978**

*Cognition and categorization* / ed. Eleanor Rosch, Barbara B. Lloyd. – Hillsdale : Lawrence Erlbaum Associates, 1978

**Concepts 1987**

*Concepts and conceptual development : ecological and intellectual factors in categorization* / ed. Ulric Neisser. – Cambridge : Cambridge University Press, 1987

**Croft 1986**

Croft W. Bruce: *User-specified domain knowledge for document retrieval* // Association for Computing Machinery (ACM) Conference on Research and Development in Information Retrieval, 1986 September 8-10 Pisa, Italy / ed. F. Rabitti. – Baltimore : Association for Computing Machinery, 1986 s. 201-206

**Cruse 1986**

Cruse D. A.: *Lexical semantics*. – Cambridge : Cambridge University Press, 1986

**Ćwiekowa 1988**

Ćwiekowa Jadwiga: *Opracowanie tematyczne piśmiennictwa : język haseł przedmiotowych, katalog przedmiotowy*. – Warszawa : Państwowe Wydawnictwo Naukowe, 1988

**Ćwiekowa 1974**

Ćwiekowa Jadwiga: *Opracowanie przedmiotowe piśmiennictwa : problemy teoretyczne i metodyczne*. – Warszawa : Ośrodek Informacji Naukowej PAN, 1974

**Daniels 1986**

Daniels Penny J.: *Cognitive models in information retrieval : an evaluative review* // Journal of Documentation Vol. 42 (1986) nr 4 s. 272-304

**Dao 1989**

Dao Thi Qui: *Metodyka budowy języka informacyjno-wyszukiwawczego dla systemu informacji specjalistycznej : na przykładzie informacji o technologii preparatów enzymatycznych* // Zagadnienia Informacji Naukowej (1989) nr 2 s. 119-154

**Dervin 1986**

Dervin Brenda, Nilan Michael: *Information needs and uses* // Annual Review of Information Science and Technology Vol. 21 (1986) s. 3-33

**Dervin 1983**

Dervin Brenda: *Information as a user construct : the relevance of perceived information needs to synthesis and interpretation* // Knowledge structures and use : implications for synthesis and interpretation / ed. S. A. Wars, L. J. Reed. – Philadelphia : Temple University Press, 1983 s. 153-184

**Dervin 1980**

Dervin Brenda: *Information gaps and inequities : moving toward a reconceptualization* // Progress in Communication Sciences Vol. 2 (1980) nr 2 s. 73-112

**Dervin 1977**

Dervin Brenda: *Useful theory for librarianship : communication, not information* // Drexel Library Quarterly Vol. 13 (1977) nr 3 s. 16-32

**Dobrowolski 1956**

Dobrowolski Zygmunt: *Budowa klasyfikacji*. – Warszawa : Państwowe Wydawnictwa Techniczne, 1956

**Doroszewski 1982**

Doroszewski Witold: *Język, myślenie, działanie : rozważania językoznawcy*. – Warszawa : Państwowe Wydawnictwo Naukowe, 1982

**Ellis 1990**

Ellis David: *New horizons in information retrieval*. – London : Library Association Publishing, 1990

**Encyklopedia 1993a**

*Encyklopedia językoznawstwa ogólnego* / pod red. Kazimierza Polańskiego. – Wrocław : Ossolineum, 1993

**Encyklopedia 1993b**

*Encyklopedia kultury polskiej XX wieku. T. 2: Współczesny język polski* / pod redakcją Jerzego Bartmińskiego. – Wrocław : Wiedza o Kulturze, 1993

**Encyklopedia 1976**

*Encyklopedia współczesnego bibliotekarstwa polskiego.* – Wrocław: Ossolineum, 1976

**Filozofia 1993**

*Filozofia języka /* wybrała i wstępem opatrzyła Barbara Stanosz. – Warszawa: Aletheia, 1993

**Foskett 1996**

Foskett Anthony C.: *The subject approach to information.* – London: Library Association Publishing, 1996

**Foskett 1963**

Foskett Douglas J.: *Classification and indexing in the social sciences.* – London: Butterworths, 1963

**Francuz 1991**

Francuz Piotr: *Funkcja ilościowych cech w kategoryzacji przedmiotów.* – Lublin: Wydawnictwo Towarzystwa Naukowego Katolickiego Uniwersytetu Lubelskiego, 1991

**Frohmann 1990**

Frohmann Bernd: *Rules of indexing: critique of mentalism in information retrieval theory* // *Journal of Documentation* Vol. 46 (1990) nr 2 s. 81-101

**From classification 1997**

*From classification to "knowledge organization": Dorking revisited or "past is prelude"* / ed. by Alan Gilchrist. – Hague: International Federation for Information and Documentation, 1997

**Geeraerts 1988**

Geeraerts Dirk: *Where does prototypicality come from?* // *Topics in cognitive linguistics* / ed. by B. Rudzka-Ostyn. – Amsterdam: John Benjamins, 1988 s. 207-229

**Gołąb 1968**

Gołąb Zbigniew, Heinz Adam, Polański Kazimierz: *Słownik terminologii językoznawczej.* – Warszawa: Państwowe Wydawnictwo Naukowe, 1968

**Grabowska 1992**

Grabowska Marta: *Systemy online w bibliotekach.* – Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 1992

**Green 1992**

Green Rebeca: *Insights into classification from the cognitive sciences: ramification for index languages* // *Classification research for knowledge representation and organization: proceedings of the 5<sup>th</sup> International Study Conference on Classification Research* / ed. Nancy Williamson. – Amsterdam: Elsevier, 1992 s. 215-222



**Grochowski 1993**

Grochowski Maciej: *Konwencje semantyczne a definiowanie wyrażen językowych*. – Warszawa : Wydawnictwa Uniwersytetu Warszawskiego, 1993

**Grolier 1962**

Grolier Eric de: *A study of general categories applicable to classification and coding in documentation*. – Paris : UNESCO, 1962

**Gruber 1995**

Gruber T. R.: *Toward principles for the design of ontologies used for knowledge sharing* // International Journal of Human-Computer Studies Vol. 43 (1995) nr 5/6 s. 907-928

**Grzegorzczkova 1995**

Grzegorzczkova Renata: *Wprowadzenie do semantyki językoznawczej*. – Warszawa : Wydawnictwo Naukowe PWN, 1995

**Guarino 1995**

Guarino N.: *Formal ontology, conceptual analysis and knowledge representation* // International Journal of Human-Computer Studies Vol. 43 (1995) nr 5/6 s. 625-640

**Guiraud 1976**

Guiraud Pierre: *Semantyka*. – Warszawa : Wiedza Powszechna, 1976

**Habrajska 1996**

Habrajska Grażyna: *Kategoryzacja a klasyfikacja – potoczne i naukowe widzenie świata : analiza badań ankietowych dotyczących kategoryzacji nazw roślin* // Językowa kategoryzacja świata / pod red. Renaty Grzegorzczkovej i Anny Pajdzińskiej. – Lublin : Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1996 s. 221-233

**Harrell 1985**

Harrell Gail, Harrell E. G.: *The classification and organisation of adult fiction in large American public library* // Public Libraries Vol. 24 (1985) nr 1 s. 13-14

**Heinz 1983**

Heinz Adam: *Dzieje językoznawstwa w zarysie*. – Warszawa : Państwowe Wydawnictwo Naukowe, 1983

**Hjørland 1998**

Hjørland Birger: *Information retrieval, text composition, and semantics* // Knowledge Organization Vol. 25 (1998) nr 1/2 s. 16-31

**Hjørland 1992**

Hjørland Birger: *The concept of 'subject' in information science* // Journal of Documentation Vol. 48 (1992) nr 2 s. 172-200

**Ingwersen 1987**

Ingwersen Peter: *Towards a new research paradigm in information retrieval // Knowledge engineering : expert systems and information retrieval / ed. Irene Wormell.* – London : Taylor Graham, 1987 s. 150-189

**Ingwersen 1984**

Ingwersen Peter, De Mey Mark: *Psychological aspects of information retrieval // Social Science Information Studies Vol. 4 (1984) nr 2/3 s. 83-85*

**Ingwersen 1982**

Ingwersen Peter: *Search procedures in the library : analysed from the cognitive point of view // Journal of Documentation Vol. 38 (1982) nr 3 s. 165-191*

**International 1997a**

*International encyclopedia of information and library science / ed. by John Feather and Paul Sturges.* – London : Routledge, 1997

**International 1997b**

*International Journal of Human-Computer Studies Vol. 46 (1997) nr 2/3 passim*

**Iyer 1995**

Iyer Hemalata: *Classificatory structures : concepts, relations and representation.* – Frankfurt/Main : Indeks, 1995

**Jacob 1991**

Jacob Elin K.: *Classification and categorization : drawing the line // Advances in classification research : proceedings of the 2nd. ASIS SIG/CR Classification Research Workshop held at 54th ASIS Annual Meeting Washington DC on October 27, 1991 / ed. Barbara Kwasnik, Fidel Raya.* – Medford : Learned Information, 1991 s. 67-82

**Jakobson 1995**

Jakobson Roman: *On language.* – Cambridge : Harvard University Press, 1995

**Językowa 1996**

*Językowa kategoryzacja świata / pod red. Renaty Grzegorzycowej i Anny Pajdzińskiej.* – Lublin : Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1996

**KABA 1999**

*Język haseł przedmiotowych KABA. Centralna kartoteka haseł wzorcowych.* Telnet: [limba.buw.uw.edu.pl](http://limba.buw.uw.edu.pl)

**Kaczmarek 1990**

Kaczmarek Bożydar L. J.: *Mózg, język, zachowanie.* – Lublin : Uniwersytet Marii Curie-Skłodowskiej, 1994

**Kamiński 1992**

Kamiński Stanisław: *Pisma wybrane*. T. 4: *Nauka i metoda ; Pojęcie nauki i klasyfikacja nauk*. – Lublin : Wydawnictwo Towarzystwa Naukowego Katolickiego Uniwersytetu Lubelskiego, 1992

**Karamuftuoglu 1997**

Karamuftuoglu Murat: *Designing language games in Okapi* // Journal of Documentation Vol. 53 (1997) nr 1 s. 69-73

**Kartoteka 1994**

*Kartoteka wzorcowa języka KABA*. Cz. 1 : *Nazwy własne / praca zbiorowa pod red. Jadwigi Woźniak*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1994

**Katzeff 1990**

Katzeff Cecilia: *System demands on mental models for a fulltext database* // International Journal of Man-Machine Studies Vol. 32 (1990) nr 5 s. 483-509

**Kempton 1993**

Kempton Willett: *Gradacja kategorii a związki taksonomiczne : kubek jest odmianą filiżanki* // *Amerykańska antropologia kognitywna : poznanie, język, klasyfikacja i kultura / wybór i red. nauk. Michał Buchowski*. – Warszawa : Wydawnictwo Instytutu Kultury, 1993 s. 99-137

**Kempton 1981**

Kempton Willett: *The folk classification of ceramics : a study of cognitive prototypes*. – New York : Academic Press, 1981

**Knowledge 1987**

*Knowledge engineering : expert systems and information retrieval* / ed. Irene Wormell. – London : Taylor Graham, 1987

**Knowledge 1983**

*Knowledge structures and use : implications for synthesis and interpretation* / ed. S. A. Wars, L. J. Reed. – Philadelphia : Temple University Press, 1983

**Kokocińska-Kurek 1999**

Kokocińska-Kurek Stanisława: *System informacyjny dla dzieci i młodzieży o książkach, filmach i grach : studium teoretyczne*. – Łódź : Wydawnictwo Uniwersytetu Łódzkiego, 1999

**Kokocińska-Kurek 1994**

Kokocińska-Kurek Stanisława: *Warszawska szkoła teorii języków informacyjno-wyszukiwawczych* // *Przegląd Biblioteczny* R. 62 (1994) z. 3/4 s. 269-277

**Koshman 1993**

Koshman Sherry: *Categorization and classification revisited : a review of concept in library science and cognitive psychology* // *Current Studies in Librarianship* Vol. 17 (1993) nr 1/2 s. 26-42

**Kosyl 1993**

Kosyl Czesław: *Chrematonimy* // Encyklopedia kultury polskiej XX wieku. T. 2. Współczesny język polski / pod red. Jerzego Bartmińskiego. – Wrocław : Wiedza o Kulturze, 1993 s. 439-444

**Kotarbiński 1961**

Kotarbiński Tadeusz: *Elementy teorii poznania, logiki formalnej i metodologii nauk*. – Wrocław : Ossolineum, 1961

**Krapiec 1985**

Krapiec Mieczysław A.: *Język i świat realny*. – Lublin : Redakcja Wydawnictw Katolickiego Uniwersytetu Lubelskiego, 1985

**Krendl 1989**

Krendl Kathy A., Broihier Mary C., Fleetwood Cynthia: *Children and computers : do sex-related differences persist?* // Journal of Communication Vol. 39 (1989) nr 3 s. 85-93

**Kuhn 1968**

Kuhn Thomas: *Struktura rewolucji naukowych*. – Warszawa : Państwowe Wydawnictwo Naukowe, 1968

**Kurcz 1995**

Kurcz Ida: *Pamięć, uczenie się, język*. – Warszawa : Wydawnictwo Naukowe PWN, 1995

**Lachman 1979**

Lachman Roy, Lachman Janet L., Butterfield Earl C.: *Cognitive psychology and information processing : an introduction*. – Hillsdale : Lawrence Erlbaum Associates, 1979

**Lakoff 1988**

Lakoff George, Johnson Mark: *Metafory w naszym życiu*. – Warszawa : Państwowy Instytut Wydawniczy, 1988

**Lakoff 1987**

Lakoff George: *Women, fire, and dangerous things : what categories reveal about the mind*. – Chicago : University of Chicago Press, 1987

**Langacker 1995**

Langacker Ronald W.: *Wykłady z gramatyki kognitywnej*. – Lublin : Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1995

**Langacker 1987-1991**

Langacker Ronald W.: *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical prerequisites* ; Vol. 2: *Descriptive application*. – Stanford : Stanford University Press, 1987-1991

**Langridge 1995**

Langridge D. W.: *Classification : its kinds, elements, systems and applications*. – London : Bowker-Saur, 1992

**LCSH 1998**

*Library of Congress Subject Headings*. -21st. edition. T. 1-5. – Washington : Library of Congress, 1998

**Leśniak 1990**

Leśniak Kazimierz: *Wstęp* [do Kategorii Arystotelesa] // *Dzieła wszystkie / Arystoteles*. T. 1. – Warszawa : Państwowe Wydawnictwo Naukowe, 1990 s. 26-31

**Licklider 1970**

Licklider Joseph C.: *Biblioteki przyszłości*. – Warszawa : Państwowe Wydawnictwo Naukowe, 1970

**Liebscher 1988**

Liebscher Peter, Marchionini Gary: *Browse and analytical search strategies in full-text CD-ROM encyclopedia* // *School Library Media Quarterly* Vol. 16 (1988) nr 4 s. 223-233

**Lindsay 1984**

Lindsay Peter H., Norman Donald A.: *Procesy przetwarzania informacji u człowieka : wprowadzenie do psychologii*. – Warszawa: Państwowe Wydawnictwo Naukowe, 1984

**Loose 1993**

Loose R. M.: *Seven fundamental questions for the science of library classification* // *Knowledge Organization* Vol. 20 (1993) nr 2 s. 65-70

**Lundeen 1982**

Lundeen Gerald: *Microprocessors in personal information systems* // *Special Libraries* Vol. 72 (1981) nr 2 s. 127-137

**Lyons 1984-1989**

Lyons John: *Semantyka*. T. 1-2. – Warszawa : Państwowe Wydawnictwo Naukowe, 1984-1989

**Łozowski 1994**

Łozowski Przemysław: *Czym jest "prototyp" w semantyce diachronicznej* // *Podstawy gramatyki kognitywnej / pod red. Henryka Kardeli*. – Warszawa : Wydawnictwa Uniwersytetu Warszawskiego, 1994 s. 243-254

**Łysakowski 1946**

Łysakowski Adam: *Katalog przedmiotowy : podręcznik*. – Warszawa: Państwowe Zakłady Wydawnictw Szkolnych, 1946

**Łysakowski 1928**

Łysakowski Adam: *Katalog przedmiotowy*. Cz. 1 : *Teoria*. – Wilno : Wydawnictwa Uniwersyteckiej Biblioteki Publicznej, 1928

**MacNamara 1993**

MacNamara John: *Logika i psychologia : rozważania z pogranicza nauk*. – Warszawa : Wydawnictwo Naukowe PWN, 1993

**Maćkiewicz 1990**

Maćkiewicz Jolanta: *Kategoryzacja a językowy obraz świata // Językowy obraz świata / red. Jerzy Bartmiński. – Lublin : Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1990 s. 42-67*

**Maćkiewicz 1983**

Maćkiewicz Jolanta: *Świat widziany poprzez język // Gdańskie Zeszyty Humanistyczne R. 26 (1983) nr 30 s. 131-150*

**Mały słownik 1994**

*Mały słownik etyczny / pod red. naukową Stanisława Jedynaka. – Bydgoszcz : Branta, 1994*

**Marciszewski 1998**

Marciszewski Witold: *Sztuczna inteligencja. – Kraków : Znak, 1998*

**Marciszewski 1977**

Marciszewski Witold: *Metody analizy tekstu naukowego. – Warszawa : Państwowe Wydawnictwo Naukowe, 1977*

**Marciszewski 1973**

Marciszewski Witold: *Relacje tematyczno-hierarchiczne w językach deskryptorowych // Zagadnienia Informacji Naukowej (1973) nr 2 s. 35-51*

**Maron 1977**

Maron M. E.: *On indexing, retrieval and the meaning of about // Journal of the American Society for Information Science Vol. 28 (1977) nr 1 s. 38-43*

**McArthur 1988**

McArthur Tom: *Worlds of reference : lexicography, learning and language from the clay tablet to the computer. – Cambridge : Cambridge University Press, 1988*

**McClellan 1981**

McClellan A. W.: *The reading demension in effectiveness and services // Library Review Vol. 30 (1981) s. 77-86*

**Medin 1989**

Medin Douglas L.: *Concepts and conceptual structur // American Psychologist (1989) nr 44 s. 1469-1481*

**Medin 1984**

Medin Douglas L., Smith Edward E.: *Concepts and concept formation // Annual Review of Psychology (1984) nr 35 s. 113-138*

**Miller 1980**

Miller Georgij Il'ič: *Metodologičeskie problemy predmetizacii : sravnitel'nyj analiz principov postroenija predmetnyh rubrik v SSSR i SŠA. – Moskva : Nauka, 1980*

**Mills 1960**

Mills J.: *A modern outline of library classification. – London : Chapman and Hall, 1960*

**Miodunka 1980**

Miodunka Władysław: *Teoria pól językowych : społeczne i indywidualne ich uwarunkowania.* – Kraków : Państwowe Wydawnictwo Naukowe, 1980

**Nauka 1996**

*Nauka o języku dla polonistów /* pod red. Stanisława Dubisza. – Warszawa : Książka i Wiedza, 1996

**Nowakowska-Kempna 1996**

Nowakowska-Kempna Iwona: *Pies i jego konceptualizacja w kontekście formułowania ocen // Opuscula Polonica et Russica T. 4 (1996) s. 53-59*

**O'Connor 1993**

O'Connor B. C.: *Browsing : framework for seeking functional information // Knowledge Vol. 15 (1993) nr 2 s. 211-232*

**Oddy 1977**

Oddy Robert N.: *Information retrieval through man-machine dialogue // Journal of Documentation Vol. 33 (1977) nr 1 s. 1-14*

**Ofori-Dwumfuo 1984**

Ofori-Dwumfuo G. O.: *Using a cognitive model of dialogue for reference retrieval // Journal of Information Science Vol. 9 (1984) nr 1 s. 19-28*

**Parasuraman 1990**

Parasuraman S., Igarria M.: *An examination of gender differences in the determinants of computer anxiety and attitudes toward microcomputers among managers // International Journal of Man-Machine Studies Vol. 32 (1990) nr 3 s. 327-340*

**Pawłowski 1986**

Pawłowski Tadeusz: *Tworzenie pojęć w naukach humanistycznych.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1986

**Pejtersen 1987**

Pejtersen Annelise Mark, Olsen Svend Erik, Zunde Pranas: *Development of a term association interface for browsing bibliographic data bases based on end users' word associations // Knowledge engineering : expert systems and information retrieval / ed. Irene Wormell. – London : Taylor Graham, 1987 s. 92-112*

**Pejtersen 1983**

Pejtersen Annelise Mark, Austin Jutta: *Fiction retrieval : experimental design and evaluation of a search system based on users' value criteria // Journal of Documentation Vol. 39 (1983) nr 4 s. 230-246 ; Vol. 40 (1984) nr 1 s. 25-35*

**Pejtersen 1979**

Pejtersen Annelise Mark.: *The meaning of 'about' in fiction indexing and retrieval // ASLIB Proceedings Vol. 31 (1979) nr 5 s. 251-257*

**Pelc 1984**

Pelc Jerzy: *Wstęp do semiotyki.* – Warszawa : Wiedza Powszechna, 1984

**Picchocka 1992**

Picchocka Elżbieta: *Analiza "Słownika terminologicznego informacji naukowej" metodą ciągów definicyjnych* : praca magisterska wykonana pod kier. doc. dr hab. Bożenny Bojar. – Warszawa : Uniwersytet Warszawski. Wydział Historyczny. Instytut Bibliotekoznawstwa i Informacji Naukowej, 1992. – [Maszynopis]

**Pisarek 1967**

Pisarek Walery: *Pojęcie pola wyrazowego i jego użyteczność w badaniach stylistycznych* // Pamiętnik Literacki R. 58 (1967) z. 2 s. 493-516

**Podstawy 1994**

*Podstawy gramatyki kognitywnej* / pod red. Henryka Kardeli. – Warszawa : Wydawnictwa Uniwersytetu Warszawskiego, 1994

**Pollitt 1990**

Pollitt Arthur S.: *Intelligent interfaces to text retrieval systems* // Text retrieval : the state of art / ed. P. Gillman. – London: Taylor Graham, 1990 s. 192-208

**Pottier 1963**

Pottier Bernard: *Recherches sur l'analyse sémantique en traduction mécanique*. – Nancy, 1963

**Proceedings 1996**

*Proceedings CoLIS 2 Second International Conference on Conceptions of Library and Information Science. Integration in Perspective October 13-16, 1996* / ed. Peter Ingwersen, Niels Ole Pors. – Copenhagen : The Royal School of Librarianship, 1996

**Próchnicka 1991**

Próchnicka Maria: *Informacja a umysł*. – Kraków : Universitas, 1991

**Pszczółowski 1978**

Pszczółowski Tadeusz: *Mala encyklopedia prakseologii i teorii organizacji*. – Wrocław : Ossolineum, 1978

**Putnam 1995**

Putnam Hilary: *Words and life* / ed. James Conant. – Cambridge : Harvard University Press, 1995

**Puzynina 1997**

Puzynina Jadwiga: *Słowo – wartość – kultura*. – Lublin : Wydawnictwo Towarzystwa Naukowego Katolickiego Uniwersytetu Lubelskiego, 1997

**RAMEAU 1998**

*RAMEAU : noms communs, noms géographiques suivis de subdivisions chronologiques*. – Montpellier : Agence bibliographique de l'enseignement supérieur, 1998



**Ramsey 1983**

Ramsey H. Rudy, Grimes Jack D.: *Human factors in interactive computer dialog* // Annual Review of Information Science and Technology Vol. 18 (1983) s. 29-59

**Ranganathan 1967**

Ranganathan Shiyali Ramamrita: *Prolegomena to library classification*. – London : Asia, 1967

**Rey 1983**

Rey Georges: *Concepts and stereotypes* // Cognition (1983) nr 15 s. 237-262

**Roget 1966**

*Roget's thesaurus of English words and phrases*. – New edition completely revised, modernized and abridged / by Robert A. Dutch. – Harmondsworth : Penguin Book, 1966

**Rosch 1975**

Rosch Eleanor, Mervis Caroly B.: *Family resemblances: studies in the internal structure of categories* // Cognitive Psychology (1975) nr 7 s. 573-605

**Sadowska 1991**

Sadowska Jadwiga: *Instrukcja tematowania i katalogu przedmiotowego*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1991

**Salton 1986**

Salton Gerard: *Another look at automatic text-retrieval systems* // Communications of Association for Computing Machinery Vol. 29 (1986) nr 7 s. 648-656

**Salton 1970**

Salton Gerard: *Dynamic information and library processing*. – Englewood Cliffs : Prentice-Hall, 1975

**Shera 1966**

Shera Jesse H.: *Libraries and the organization of knowledge*. – London : Crosby Lockwood and Son, 1966

**Sitarska 1990**

Sitarska Anna: *Systemowe badanie bibliotek : studium metodologiczne*. – Łódź : Wydawnictwo Uniwersytetu Łódzkiego, 1990

**SJPD**

*Słownik języka polskiego* / pod red. Witolda Doroszewskiego. T. 1-4. – Warszawa : Wiedza Powszechna, 1958-1962 ; T. 5-11. – Warszawa : Państwowe Wydawnictwo Naukowe, 1963-1969

**SJPS**

*Słownik języka polskiego* / red. naukowy Mieczysław Szymczak. – Wyd. 5. T. 1-3. – Warszawa : Państwowe Wydawnictwo Naukowe, 1988

**Słownik 1997**

*Słownik literatury popularnej* / pod red. Tadeusza Żabskiego. – Wrocław : Towarzystwo Przejściół Polonistyki Wrocławskiej, 1997

**Słownik 1996**

*Słownik pojęć filozoficznych* / pod red. Władysława Krajewskiego. – Warszawa : Scholar, 1996

**Słownik 1993a**

*Słownik encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych* / red. nauk. Bożenna Bojar. – Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 1993

**Słownik 1993b**

*Słownik języka haseł przedmiotowych Biblioteki Narodowej : stan na dzień 31 grudnia 1992 r.* / oprac. Janina Trzcńska i Ewa Stępiakowa. – Wyd. 2 popr. i rozszerz. T. 1-2. – Warszawa : Biblioteka Narodowa, 1993

**Smith 1981**

Smith Edward E., Medin Douglas L.: *Categories and concepts*. – Cambridge : Harvard University Press, 1981

**Smith 1987**

Smith Linda C.: *Artificial intelligence and information retrieval* // Annual Review of Information Science and Technology Vol. 22 (1987) s. 41-77

**Soergel 1985**

Soergel Dagobert: *Organizing information : principles of database and retrieval systems*. – London : Academic Press, 1985

**Soper 1976**

Soper Mary E.: *Characteristics and use of personal collections* // Library Quarterly Vol. 46 (1976) nr 4 s. 397-415

**Sosińska 1999**

Sosińska-Kalata Barbara: *Modele organizacji wiedzy w systemach wyszukiwania informacji o dokumentach*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1999

**Sosińska 1995**

Sosińska-Kalata Barbara: *Podręcznik UKD dla bibliotekarzy i pracowników informacji*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1995

**Sosińska 1993**

Sosińska-Kalata Barbara: *Uniwersalna Klasyfikacja Dziesiętna : podręcznik*. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1993

**Sosińska 1988**

Sosińska Barbara: *Relacje między planem treści i planem wyrażania w językach informacyjno-wyszukiwawczych* : rozprawa doktorska wykonana pod kier. doc. dr hab. Bożenny Bojar. – Warszawa : Uniwersytet Warszawski. Wydział Neofilologii, 1988. – [Maszynopis]

**Sowa 1995**

Sowa John F.: *Top-level ontological categories* // International Journal of Human-Computer Studies Vol. 43 (1995) nr 5/6 s.669-685

**Spiller 1980**

Spiller David: *The provision of fiction for public libraries* // Journal of Librarianship Vol. 12 (1980) nr 4 s. 238-264

**Szrejder 1975**

Szrejder Jurij A.: *Równość, podobieństwo, porządek*. – Warszawa : Wydawnictwa Naukowo-Techniczne, 1975

**Ścibor 1996**

Ścibor Eugeniusz: *Klasyfikacja piśmiennictwa*. – Olsztyn : Wydawnictwo Wyższej Szkoły Pedagogicznej, 1996

**Ścibor 1995**

Ścibor Eugeniusz: *Rewizja UKD w dziedzinie języków, językoznawstwa i literatury pięknej* // Zagadnienia Informatyki (1995) nr 1/2 s. 19-34

**Ścibor 1982**

Ścibor Eugeniusz: *Typologia strukturalna języków informacyjnych : rozprawa habilitacyjna*. – Warszawa : Instytut Informatyki, Technicznej i Ekonomicznej, 1982

**Tabakowska 1995**

Tabakowska Elżbieta: *Gramatyka i obrazowanie : wprowadzenie do językoznawstwa kognitywnego*. – Kraków : Wydawnictwo Oddziału Polskiej Akademii Nauk, 1995

**Tadeusiewicz 1998**

Tadeusiewicz Ryszard: *Metody komputerowego wspomaganie nauczania jako jeden z wyróżników społeczeństwa informacyjnego* // Materiały na Międzynarodową Konferencję nt.: Współpraca bibliotek naukowych w zakresie automatyzacji, Kraków, 16-19.11.1998 / [red. t. Maria Burchard, Maria Lenartowicz]. – Warszawa : Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 1998 s. 9-20

**Taylor 1991**

Taylor John R.: *Linguistic categorization : prototypes in linguistic theory*. – Oxford : Clarendon Press, 1991

**Taylor 1986**

Taylor Robert S.: *Value-added processes in information systems*. – Norwood : Ablex, 1986

**Taylor 1985**

Taylor Robert S.: *Information values in decision contexts* // Information Management Review Vol. 1 (1985) nr 1 s. 47-55

**Taylor 1984**

Taylor Robert S.: *Value-added processes in document-based systems : abstracting and indexing services* // Information Services and Use Vol. 4 (1984) nr 3 s. 127-146

**Taylor 1968**

Taylor Robert S.: *Question-negotiation and information seeking in libraries* // College and Research Libraries Vol. 29 (1968) nr 3 s. 178-194

**Text 1990**

*Text retrieval : the state of art* / ed. P. Gillman. – London: Taylor Graham, 1990

**Temple 1989**

Temple L., Lips H. M.: *Gender differences and similarities in attitudes toward computers* // Computers in Human Behaviour Vol. 5 (1989) nr 4 s. 215-226

**Tezaurus 1996**

*Tezaurus technologii żywności i przemysłu spożywczego.* – Warszawa : Centralna Biblioteka Rolnicza, 1996

**Tezaurus 1985**

*Tezaurus z zakresu poczty* / [oprac.] Czesław Bedronek [i in.]. – Warszawa : Centralny Ośrodek Badawczo-Rozwojowy Poczty, 1985

**Tokarski 1996**

Tokarski Ryszard: *Ramy interpretacyjne a problemy kategoryzacji* // Językowa kategoryzacja świata / pod red. Renaty Grzegorzyczykowej i Anny Pajdzińskiej. – Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, 1996 s. 97-112

**Tokarski 1993**

Tokarski Ryszard: *Słownictwo jako interpretacja świata* // Encyklopedia kultury polskiej XX wieku. T.2. Współczesny język polski / pod red. J. Bartmińskiego. – Wrocław : Wiedza o Kulturze, 1993 s. 335-362

**Tokarski 1984**

Tokarski Ryszard: *Struktura pola znaczeniowego : studium językoznawcze.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1984

**Tokarz 1985**

Tokarz Aleksandra: *Rola motywacji poznawczej w aktywności twórczej.* – Wrocław : Ossolineum, 1985

**Trzebiński 1981**

Trzebiński Jerzy: *Twórczość a struktura pojęć.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1981

**UKD 1997**

*Uniwersalna Klasyfikacja Dziesiąta.* Publikacja nr UDC – P022 autoryzowana przez Konsorcjum UKD nr licencji UDC – 9709 / [tablice oprac. Lucyna Anna Bielicka, Adam Stopa, Teresa Turowska]. – Wyd.

skrócone dla bieżącej bibliografii narodowej i bibliotek publicznych. –  
Warszawa : Biblioteka Narodowa, 1997

**UKD 1995**

*Uniwersalna Klasyfikacja Dziesiąta FID 607. Tablica 1c: Poddziały wspólne języka. Dział 8: Filologia Językoznawstwo i języki ; Literatura piękna. – Wyd. 2 pełne w jęz. pol. aktualne na dzień 1. VII. 1979 r. wg „Extensions and Corrections to the UDC” Seria 17. – Warszawa : Instytut Informacji Naukowej, Technicznej i Ekonomicznej, 1995*

**Ungurian 1983**

Ungurian Olgierd: *Kategoria semantyczna wymiaru "Przedmiot dokumentu"* // *Zagadnienia Informacji Naukowej* (1983) nr 1 s. 29-54

**Ungurian 1982**

Ungurian Olgierd: *Ogólna struktura języka informacyjnego : propozycja "szkieletu organizacyjnego" słownictwa* // *Zagadnienia Informacji Naukowej* (1982) nr 2 s. 21-40

**Ungurian 1976**

Ungurian Olgierd: *Elementy teorii języków informacyjnych.* – Warszawa : Ośrodek Informacji Naukowej PAN, 1976

**Ungurian 1975**

Ungurian Olgierd: *Teoria i praktyka klasyfikacji fasetowej S. R. Ranganathana.* – Warszawa : Centrum Informacji Naukowej, Technicznej i Ekonomicznej, 1975

**Uschold 1996**

Uschold M., Gruninger M.: *Ontologies : principles, methods and applications* // *Knowledge Engineering Review* Vol. 11 (1996) nr 2 s. 93-136

**Vickery 1997**

Vickery Brian C.: *Ontologies* // *Journal of Information Science* Vol. 23 (1997) nr 4 s. 277-286

**Vickery 1987a**

Vickery Alina, Vickery Brian C.: *Information science : theory and practice.* – London : Bowker-Saur, 1987

**Vickery 1987b**

Vickery Alina, Brooks Helen M.: *PLEXUS : the expert system for referral* // *Information Processing and Management* Vol. 23 (1987) nr 2 s. 99-117

**Vickery 1987c**

Vickery Alina, Brooks Helen M.: *A reference and referral system using expert system techniques* // *Journal of Documentation* Vol. 43 (1987) nr 1 s. 1-23

**Vickery 1975**

Vickery Brian C.: *Classification and indexing in science and technology.* – London : Butterworths, 1975

**Vickery 1953**

Vickery Brian C.: *Systematic subject indexing* // Journal of Documentation Vol. 9 (1953) nr 1 s. 51-62

**Vygotskij 1971**

Vygotskij Lev S.: *Wybrane prace psychologiczne.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1971

**Walker 1958**

Walker R. S.: *Problem child : some observations on fiction, with a sketch of a new system of classification* // Librarian and Book World Vol. 47 (1958) nr 1 s. 21-28

**Wereszczyńska-Cisło 1997a**

Wereszczyńska-Cisło Barbara: *Charakterystyka i możliwości eksploatacji hipertekstu* // Zagadnienia Informacji Naukowej (1997) nr 2 s. 17-30

**Wereszczyńska-Cisło 1997b**

Wereszczyńska-Cisło Barbara: *Wykorzystanie relacji skojarzeniowych specyfikowanych w teaurusie do określania hipertekstowych powiązań* // Zagadnienia Informacji Naukowej (1997) nr 2 s. 31-46

**Wereszczyńska-Cisło 1986a**

Wereszczyńska-Cisło Barbara, Ogórkiewicz Wiesława: *Wpływ szczegółowej specyfikacji relacji na efektywność wyszukiwania informacji z zakresu technologii żywności.* – Warszawa : Centrum Informacji Naukowej, Technicznej i Ekonomicznej, 1986

**Wereszczyńska-Cisło 1986b**

Wereszczyńska-Cisło Barbara, Ogórkiewicz Wiesława: *Pytania informacyjne w ujęciu teoretycznym* // Zagadnienia Informacji Naukowej (1986) nr 1 s. 41-59

**Wereszczyńska-Cisło 1984**

Wereszczyńska-Cisło Barbara: *Przydatność specyfikacji relacji kojarzeniowych w procesie wyszukiwania informacji z zakresu wybranych gałęzi technologii żywności* : rozprawa doktorska wykonana pod kier. prof. dr hab. Witolda Marciszewskiego. – Warszawa : Uniwersytet Warszawski. Wydział Neofilologii, 1984. – [Maszynopis]

**Wierzbicka 1999**

Wierzbicka Anna: *Język, umysł, kultura* / wybór prac pod redakcją Jerzego Bartmińskiego. – Warszawa : Wydawnictwo Naukowe PWN, 1999

**Wierzbicka 1996**

Wierzbicka Anna: *Semantics : primes and universals.* – Oxford : Oxford University Press, 1996

**Willetts 1975**

Willetts Margaret: *An investigation of the nature of the relation between terms in thesauri* // Journal of Documentation Vol. 31 (1975) nr 3 s. 158-184

**Wilson 1968**

Wilson Patric: *Two kinds of power : an essay on bibliographic control.* – Berkeley : University of California Press, 1968

**Wilson 1984**

Wilson Thomas D.: *The cognitive approach to information-seeking behaviour and information use* // Social Science Information Studies Vol. 4 (1984) nr 4 s. 197-204

**Wilson 1981**

Wilson Thomas D.: *On user studies and information needs* // Journal of Documentation Vol. 37 (1981) nr 1 s. 3-15

**Wittgenstein 1972**

Wittgenstein Ludwig: *Dociekania filozoficzne.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1972

**Wojtasiewicz 1992**

Wojtasiewicz Olgierd: *Wstęp do teorii tłumaczenia.* – Warszawa : TEPIS, 1992

**Wormell 1984**

Wormell Irene: *Cognitive aspects in natural language and free-text searching* // Social Science Information Studies (1984) nr 4 s. 131-141

**Woźniak 1999**

Woźniak Jadwiga: *Słownik języka informacyjno-wyszukiwawczego jako słownik przekładowy* // Semiosis lexicographica T. 3 (1999) s. 43-50

**Woźniak 1998**

Woźniak Jadwiga: *Cechy dobrego języka informacyjno-wyszukiwawczego* // Zagadnienia Informacji Naukowej (1998) nr 2 s. 3-17

**Woźniak 1997**

Woźniak Jadwiga: *Kognitywizm w informacji* // Zagadnienia Informacji Naukowej (1997) nr 2 s. 3-16

**Woźniak 1989**

Woźniak Jadwiga: *O tak zwanych potrzebach informacyjnych* // Zagadnienia Informacji Naukowej (1989) nr 1 s. 39-59

**Wójcik 1969**

Wójcik Tadeusz: *Prakseosemiotyka : zarys teorii optymalnego znaku.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1969

**Wójcik 1965**

Wójcik Tadeusz: *Zarys teorii klasyfikacji : zagadnienia formalne.* – Warszawa : Państwowe Wydawnictwo Naukowe, 1965

**Zunde 1969**

Zunde Pranas, Dexter M. E.: *Indexing consistency and quality* // American Documentation Vol. 20 (1969) nr 3 s. 259-264

## SUMMARY

The main subject of this work is the use of lexical categorisation in the indexing languages used in information retrieval systems for documents. My intention was not to create a specific categorisation but to characterise the problem of categorisation with insight and from different points of view. Nevertheless, this work is only an introduction to the vast problems connected with the issue of categorisation in indexing languages. It sets in order the current state of knowledge providing grounds for the need and possibilities for further research. I tried also to suggest a new way of perceiving information-retrieval systems and languages.

The book consists of three chapters. The first one: *Language and document representation in information retrieval systems* is about language and the representation of documents. It covers: the concept of subject; definitions and ways of understanding the term of meaning; the importance of definitions, explications and examples in a process of learning and usage of indexing languages; the problem of document adequacy; the necessity of further research in a range of pragmatism of indexing languages; further relations between the design of indexing languages and the information structure of the documents likely to be indexed; the state of cognitive research in information science. The description of conceptual organisation and retrieval of text by the American historians as an example of cognitive approach to information categorisation and retrieval ends the chapter.

The second chapter: *Categorisation: a selective review of research* is an attempt to create the background for categorisation in indexing languages. It starts with the terminology in categorisation and categories. The main section presents selected general outline for categorisation beginning with the Aristotelian model of categorisation, through the classical approach to categorisation in semantics and the Wittgenstein concept of family resemblance and concluding with prototype categories.

The third chapter: *Categorisation in the theory and practice of indexing languages* is about categorisation in the theory and practice of indexing languages. The three subtitles are the equivalent of the most popular ways of understanding the subject of the categorisation in information science publications, e.g.: as a system of categories (*Categories in indexing languages*); as a process of forming the system of categories (*The development*



of *category-based indexing languages*) and as a process of categorisation of the new objects into categories already existed in an existing language (*Categorisation of objects*).

*Categories in indexing languages.* There were four indexing languages taken into consideration to analyse the system of categories: Universal Decimal Classification (*Categories in the Universal Decimal Classification*), Adam Lysakowski and KABA subject headings systems (*Categories in subject headings systems*), and the Thesaurus of Food Technology and Food Industry (*Categories in the Thesaurus of Food Technology and Food Industry*). Also categories in AltaVista (*Categories in the Internet search engines*) were analysed.

*The development of category-based indexing languages.* The process of forming the system of categories was analysed from the following points of view: the semantic fields theory, the theory of indexing languages and the concept of ontology which has its roots in an artificial intelligence and knowledge engineering.

*Categorisation of objects.* The process of categorisation was characterised by highlighting the important research and practical problems in this area and their implications. To find the solutions requires further separate research. The advisability and justification of those are explained at the end.

The conclusions close the discussion and provide recommendations for further categorisation research in the theory and practice of indexing languages. The general conclusions are:

- It would be beneficial for indexing languages to enrich the traditional connotation-denotation conception of meaning with Wittgenstein's conception of meaning as a use of words.
- The context of information is an important element of document representation and would greatly benefit the effectiveness of document representation.
- Further category research should be taken in relation to the quality of document representation and to categorisation in indexing languages.

One of the requirements for the quality of document representation is the existence of relevant semantic classes in indexing languages. The category approach gives the opportunity to point those semantic features of indexing languages. Connotation-denotation approach does not create such a wide opportunity.

**SERIA WYDAWANA  
Z INICJATYWY  
INSTYTUTU INFORMACJI NAUKOWEJ  
I STUDIÓW BIBLIOLOGICZNYCH  
UNIwersYTETU WARSZAWSKIEGO  
ORAZ WYDAWNICTWA SBP**