

Impresjony słów. Język naturalnych i sztucznych sieci neuronowych

Abstract: It was shown that mental representations of objects created in human minds during the learning process, which take the form of hierarchically connected neurons called semblions, have properties explaining the hierarchical structure of metaphors and other mind tools embodied in Lakoff's cognitive linguistics concept. It was proven that, at the neurological level, semblions of objects, concepts, ideas and models can be associated with neuronal representations of phonemes heard in coincidence with other objects, creating new semblions which correspond to words. It was shown how activation of semblions can result in recalling, associating, thinking and other higher mental functions which are necessary to use natural language. It was considered why semblions representing mathematical notions enable an adequate description of numerous phenomena of the physical world but, at the same time, their polymodal counterparts can be used to describe *qualia*. The complementarity between Horzyk's associative intelligence model and Galus' model of architecture of self-aware systems was pointed out, which can be used to create artificial self-aware systems able to use both natural language and formal languages with understanding.

Keywords: consciousness, self-awareness, semblion, the drive of understanding, model of the mind, synaptic fields, synaptic coupling, attention switching

1. Wstęp

Celem pracy jest wykazanie, że niedawno odkryte rozległe struktury neuronowe, stanowiące reprezentacje neuronalne postrzeganych obiektów, mogą być podłożem biofizycznym zapamiętywania wrażeń o różnym stopniu ogólności, od rejestracji prostych bodźców, poprzez subiektywne wrażenia zmysłowe (tak zwane *qualia*), aż po złożone obrazy sceny oraz ogólne pojęcia, idee i modele rzeczywistości. Te zapamiętywane pojęcia i wrażenia mogą być kojarzone z zestawami znaków dźwiękowych lub obrazowych, tworząc język komunikacji. Hipoteza ta umożliwi rozważenie, czy zaproponowane ostatnio samouczące się struktury równoległego przetwarzania zawartości pamięci pozwalają zaprojektować sztuczne systemy kognitywne, zdolne do tworzenia języka i rozumnego posługiwania się nim w środowisku innych systemów autonomicznych, nie wykluczając ludzi.

Naturalny, rozwinięty język komunikacji jest jedną z umiejętności, która najbardziej wyróżnia ludzi spośród wszelkich innych istot żywych. Dotyczy to również, i to w jeszcze większym stopniu, języków formalnych, w tym matematyki. Za pomocą języka możemy opisać najbardziej subtelne stany emocjonalne, niezwykle wrażenia subiektywne, skomplikowane zależności logiczne, niezmiernie złożone struktury, konstrukcje materialne i umysłowe, a także precyzyjne, ściśle zależności struktur matematycznych. Zadziwiająca jest elastyczność języka, pozwalająca na przeniesienie naszych odczuć i wyobrażeń na innych ludzi, dysponujących tym samym językiem. Jeszcze bardziej interesująca jest zgodność tworzonych w naszych umysłach struktur matematycznych opisujących modele zjawisk z potwierdzanymi doświadczalnie prawami przyrody i rzeczywistym przebiegiem tych zjawisk. Postęp w rozumieniu zasad funkcjonowania naszego umysłu pozwala nam rozważać hipotezy, jakie procesy mózgowie odpowiadają za wytworzenie tak bogatego i skutecznego języka oraz w jaki sposób uzyskujemy zdolność do adekwatnego opisu świata, który nas otacza.

Punktem wyjścia tych rozważań będzie perspektywa językoznawstwa kognitywnego. Nie będziemy się jednak zajmować hipotezą języka myśli Fodora i Pylyshyna, komputacjonizmem ani wizją języka formalnego, rozwiniętą najpełniej przez szkołę Noama Chomsky'ego i jego kontynuatorów, wyróżniającą syntaktykę jako podstawę struktur języka i wyznaczającą ramę dla semantyki (teoria lingwistyczna [Chomsky 1982]). W świetle osiągnięć współczesnej kognitywistyki postulaty istnienia tak zwanego modułu językowego lub wrodzonej, uniwersalnej gramatyki modyfikowanej przez lokalne formy języków naturalnych posiadają ograniczoną użyteczność dla wyjaśnienia, jak umysł produkuje język.

Językoznawstwo matematyczne, odnoszące się do języków regularnych, w niewielkim stopniu wyjaśnia, jak umysł tworzy pojęcia abstrakcyjne języka naturalnego. Zwierzęta skutecznie komunikują się w języku gestów, mimiki, zapachów i dźwięków nieartykułowanych, niepoddającym się analizie logicznej, a jego gramatyka na pewno nie ma charakteru formalnego. Można natomiast wewnątrz ludzkiego mózgu poszukiwać struktur pełniących w pewnym sensie funkcję „modułu językowego” postulowanego przez Chomsky'ego, czego dokonamy w następnych rozdziałach. Interesujące jest, że proste pojęcia prematematyczne mogą być formułowane w języku naturalnym, dalekim od formalizmu. Jednakże tworzenie regularnego języka formalnego wymaga przetwarzania języka na wysokim stopniu uogólnienia zgodnie z zasadami logiki dostępnymi tylko dla umysłów ludzkich. W rozdziale 5 rozważona zostanie możliwość zastosowania modelu informatycznego do budowy sztucznego systemu świadomie posługującego się językiem naturalnym.

2. Narzędzia umysłu ucieleśnionego

Konstruktywne podejście do tworzenia pojęć języka i jego struktur gramatycznych należy identyfikować z koncepcją umysłu ucieleśnionego i rolą metafor.

Struktura pojęć i relacji lingwistycznych odzwierciedla doświadczenia motoryczne i relacje przestrzenne, których doświadczamy poprzez kontakt z otoczeniem. Podstawowe, subiektywne wrażenia zmysłowe, tzw. *qualia*, są podstawą określeń tworzących

opis elementarnych doznań i postrzegania konkretnych obiektów. Przeprowadzone na ich podstawie wyższe funkcje poznawcze, abstrahowanie, kategoryzacja, generalizacja, asocjacje i dalsza idealizacja, kształtują pojęcia abstrakcyjne i złożone modele oraz odpowiadające im konstrukty lingwistyczne. Proces ten opisany jest szczegółowo w pracach Galusa [2015a, 2015b, 2015c]. Nawet złożone pojęcia matematyczne odzwierciedlają nasze powszechne doznania związane z ruchem ciała w przestrzeni lub bezpośrednią obserwacją obiektów. Dotyczy to pojęć klasy lub zbioru abstrahowanych z powszechnych doświadczeń operowania zbiorami przedmiotów zebranych w ograniczonej przestrzeni, pojęcia pochodnej funkcji wywodzącej się z obserwacji rzeczywistych, dynamicznie zmiennych poruszeń ciała, jak też stopniowego zbliżania się do granicy obszaru lub mety kończącej ruch. Także pojęcie rekursji wywodzi się z doświadczeń powtarzalnych zdarzeń i celowo organizowanych działań [Lakoff i Núñez, 2000].

Narzędziami umysłu ucieleśnionego w kształtowaniu języka są, według Lakoffa i kontynuatorów tradycji kognitywistycznej, schematy wyobrażeniowe (*image schemas*), system aspektualny i metafory pojęciowe. Schematy wyobrażeniowe łączą funkcje poznawcze, perceptualną i konceptualną, a więc język i rozumowanie z postrzeganiem. Autorzy przywołują tu powszechnie wykorzystywane schematy: góra-dół, część-całość, plan-tło oraz schemat pojemnika. W opisie językowym struktura zdarzeń określana jest przez aspekt. Biorąc pod uwagę, że wszelkie procesy motoryczne posiadają ogólną strukturę odzwierciedlającą się w języku opisu, możemy go rozpatrywać jako system aspektualny [Lakoff i Johnson, 1999].

Lingwistyka kognitywna, uzasadniając ideę umysłu ucieleśnionego, duże znaczenie przywiązuje do roli, jaką w ludzkich mózgach odgrywiają neurony lustrane. Vittorio Gallese i Alvin Goldman uważają, że to właśnie neurony lustrane są podstawą ważnego mechanizmu „symulacji ucieleśnionej” [Gallese i Goldman, 1998]. Gallese określa symulacje zachodzące w mózgu jako „(...) automatyczny, nieświadomy, przedrefleksyjny mechanizm funkcjonalny prowadzący do modelowania obiektów, osób i zdarzeń. Bierze on udział w tworzeniu reprezentacji umysłowych, jest więc podstawą poznawania rzeczywistości” [Gallese, 2009, s. 196]. Symulacja w tym znaczeniu nie jest tylko naśladownictwem obserwowanych zachowań motorycznych, ale jest też procesem umysłowym, gdzie pobudzenie neuronów lustranych generuje procesy myślowe podobne do procesów zachodzących w umyśle obserwowanego osobnika. Dzięki temu możemy odgadywać zamiary i rozumieć intencje partnerów w relacjach społecznych. Jest to podstawą empatii, czyli zdolności do wyobrażenia sobie bólu i radości, i wszelkich innych stanów psychicznych, rozumienia przyczyn tych stanów, które i nam są bliskie. Może to rodzić chęć wymiany informacji o tych stanach, co sprzyja formowaniu języka. Zdolność do empatii łączona z instynktem macierzyńskim prowadzi według Fransa de Waala do altruizmu, tym szerszego, im szerszy jest pogląd na świat i im większa świadomość własnej roli w tym świecie. Prowadzi to do współpracy i uczenia jako procesu uspołecznionego, co także sprzyja rozwojowi języka [de Waal, 2014]. Według Vilayanura Subramaniana Ramachandrana struktury, w których występują neurony lustrane, odgrywają istotną rolę w tworzeniu reprezentacji abstrakcyjnych. Ze względu na ich polimodalny charakter mogą one odgrywać fundamentalną rolę w tworzeniu metafor [Ramachandran, 2012].

Zarówno sam Lakoff, jak i jego kontynuatorzy zwrócili uwagę, że akt rozumienia języka oznacza zwykle znalezienie w umyśle znanego odpowiednika obiektu lub procesu opisanego słownie. Można to interpretować jako znalezienie metafory odpowiadającej tym obiektom. Z tymi odpowiednikami obiektów lub procesów musieliśmy się uprzednio zapoznać w toku życiowych doświadczeń. Tak więc rozumiemy wypowiedzi poprzez odnajdywanie w naszej pamięci metafor im odpowiadających, podobnie jak rozpoznajemy obiekty i procesy poprzez ich porównanie ze wzorcami zagnieżdżonymi w naszym umyśle. George Lakoff wraz z Markiem Johnsonem dowodzą, że: „(...) cały system pojęciowy, w ramach którego myślimy i działamy, jest z natury metaforyczny” [Lakoff i Johnson, 2010]. Metafory można charakteryzować jako rozumienie i doświadczanie jakiejś rzeczy w kategoriach innej rzeczy lub ogólnie transformowanie obiektów jednej dziedziny w obiekty innej dziedziny. Taka definicja obejmuje zarówno poziom doświadczeń cielesnych, jak i poziom ekspresji językowej. Przy czym mniej istotne jest tu podobieństwo zbiorów obiektów, a bardziej podobieństwo struktury relacji między nimi. Przenoszenie gotowych struktur na nowe dziedziny poznania pozwala na podniesienie poziomu abstrakcji w procesie poznania.

Najwyższe poziomy abstrakcji dotyczą zagadnień matematyki i logiki formalnej. Dla wyjaśnienia, jak teoria metafor może tworzyć język matematyki, kognitywistyka, obok umysłu ucieleśnionego, wprowadza pojęcie matematyki ucieleśnionej. Ucieleśnienie zdolności matematycznych oznacza, że wyrastają one z naszych odzwierzęcych, prymitywnych kompetencji numerycznych, takich jak estymacja, subitacja, rudymen tarne liczenie, dodawanie i odejmowanie, ale także z codziennych doświadczeń obejmujących percepcję relacji przestrzennych, statycznych i dynamicznych oraz orientację własnego ciała względem sąsiednich obiektów, operacje grupowania obiektów w przestrzeni ograniczonej „pojemnikiem”, posługiwanie się zapisem symbolicznym, uporczywe powtarzanie czynności „bez końca” itp. Jednakże prawdziwa matematyka wymaga zaprzęgnięcia subtelnych metafor, których hierarchie wprowadzają Lakoff i Núñez [2000].

Trudno wyobrazić sobie wzajemną komunikację bez zrozumienia intencji interlokutora dzięki symulacji oraz bez wykorzystywania metafor. Szczególnie opis prostych wrażeń zmysłowych, tzw. *qualiów*, nie jest możliwy bez języka metafor, a nawet języka poezji. Nie sposób opisać smaku wina, zapachu róży lub brzmienia fletu, nie odwołując się do polimodalnych przenośni. Dlatego należy uznać fundamentalną rolę narzędzi umysłu ucieleśnionego w poznaniu świata i formułowaniu języka komunikacji społecznej. Bardziej złożone obiekty posiadają zwykle więcej dobrze zdefiniowanych cech i z tego powodu poddają się analizie logicznej i znacznie bardziej precyzyjnemu opisowi, o czym mówić będziemy dalej.

Oczywiście, tego rodzaju program kognitywistyczny nie znajduje uznania wśród neoplatoników, którzy twierdzą, iż język metafor nie może mieć dostatecznej precyzji formułowania analiz logicznych lub złożonych modeli matematycznych odnoszących się do fizycznej rzeczywistości czy też absolutnie abstrakcyjnych. Zwracają oni uwagę, że teoria metafor ma charakter spekulatywny, pełen nieścisłości nieakceptowalnych dla matematycznych umysłów. Kolejny zarzut dotyczy źródeł stabilności gromadzonej wiedzy i jej kumulatywnego charakteru w sensie ontogenetycznym, jak i kulturowym, który trudno jest wytłumaczyć „metaforycznym” sposobem pozyskiwania wiedzy.

Andrzej Pawelec na przykład zaznacza, że pojmowanie znaczenia postrzeganego obiektu lub rozpoznawaniego znaku nie może być efektem przywoływania jego metaforycznej formy. Krytykuje on model Lakoffa, w którym mamy niejako dwie warstwy: podstawowych struktur znaczenia wyłaniających się jakoby w „doświadczeniu bezpośrednim”, fizycznym oraz znaczeń pochodnych będących rezultatem transferów i integracji struktur elementarnych. Warstwa znaczeń bezpośrednich ma być efektem „percepcji »gestaltowej«, projektów motorycznych i szczegółowych obrazów mentalnych” [Lakoff, 1987, s. 267]. Pawelec twierdzi, że nie można zakładać, iż w obcowaniu ze światem „wyłania” się samoczynnie stabilna reprezentacja tego świata. Weźmy dla przykładu obrazy mentalne. Przywołując Piageta, Pawelec pisze: „obraz nie jest ani elementem samego myślenia, ani też bezpośrednią kontynuacją percepcji: jest to symbol przedmiotu, i to taki, który nie przejawia się jeszcze na poziomie inteligencji sensoryczno-motorycznej. Stabilna reprezentacja przedmiotu w umyśle jest efektem symbolicznego naśladownictwa tego przedmiotu, np. odróżniania go od innych przedmiotów, pokazywania go gestem czy rysowania go. »Istotna« struktura przedmiotu może się wyłonić w umyśle tylko dzięki działaniom symbolicznym...” [Pawelec, 2005, s. 258].

Ta teza stoi w wyraźnej sprzeczności z prezentowanym dalej modelem budowy hierarchicznej struktury reprezentacji mentalnych i odpowiadających im reprezentacji biofizycznych postrzeganego rzeczywistości. Począwszy od *qualiów*, aż do poziomu modelu świata, w którym ucieleśniony umysł działa.

Lakoff i Johnson [2010] ripostują, że odwzorowania metaforyczne działają systematycznie i są w stanie osiągać niezbędną precyzję, obejmując wszelkie zdarzenia, scenariusze i relacje pomiędzy obiektami wielu dziedzin. Powyższe zarzuty skutecznie obalają także Bartosz Brożek i Mateusz Hohol, uznając, że na skuteczność teorii metafor wskazują dane lingwistyczne i psychologiczne. Teoria metafor pasuje do zjawisk językowych, takich jak polisemia czy zmiana znaczenia pojęć, a zarazem dostarcza ich wyjaśnienia. W badaniach psychologicznych o charakterze behawioralnym z wykorzystaniem torowania (*priming*) wykazano, że pomimo silnego torowania metafory wciąż dominują w strukturze pojęciowej, determinując wyobrażenia i decyzje. W badaniach rozwojowych nad kształtowaniem zdolności do metaforyzacji u małych dzieci dowiedziono, że metaforyczność pojawia się w drugim etapie, kiedy dziecko uczy się stopniowo rzutować strukturę pojęciową z jednej dziedziny na drugą. Podobne efekty zaobserwowano w badaniach psychologicznych nad gestykulacją i jej porównaniu z towarzyszącym wyjaśnieniem słownym [Brożek i Hohol, 2014].

Brożek i Hohol uzupełniają narzędzia umysłu ucieleśnionego o użycie symboli, „imitację” i interakcje społeczne tworzące kulturę. Wszystkie one są ze sobą ściśle powiązane, bo przecież to symbole stanowią fundament ludzkiej kultury. Bez nich kultura i nauka nie mogłaby istnieć. Imitacja to unikatowa zdolność poznawcza umysłu, którą najdoskonalej opanował właśnie człowiek. Być może jest ona skutkiem neotenui u ludzi, ponieważ wiele zwierząt w okresie niemowlęcym i wczesnego rozwoju uczy się poprzez naśladowanie. Także niemowlęta ludzkie są prawdziwymi maszynami do imitacji zachowań rodziców. Lecz zdolność imitacji jest również osnową rozwoju kultury. Jak piszą autorzy: „(...) imitacja odgrywa zasadniczą rolę w powstaniu i ewolucji kultury”. I dalej: „(...) kumulatywną ewolucję kulturową wraz z ewolucją

matematyki (...) umożliwia konkretna zdolność poznawcza powstała przez dobór naturalny: imitacja”. Imitacja powoduje efekt „zapadki kulturowej”, kiedy kumulacja efektów rozwoju kulturowego uniemożliwia cofnięcie się do wcześniejszych stadiów. „Odziedziczone wytwory kulturowe możemy modyfikować i dodawać do nich nowe wynalazki, których z kolei, w drodze imitacji, uczą się od nas kolejne pokolenia”. Autorzy zdecydowanie opowiadają się za koncepcją umysłu i matematyki ucieleśnionej oraz teorią metafor Lakoffa. Piszą oni: „(...) koncepcja [metafor] Lakoffa najlepiej tłumaczy genezę wielu zdolności poznawczych. Wydaje się ona przy tym jedyną istniejącą teorią naukową, która systematycznie wyjaśnia, jaka jest geneza poznania matematycznego” [Brożek i Hohol, 2014].

3. Geneza poznania

Zauważmy, że mówiąc o symulacjach, metaforach, schematach wyobrażeniowych i innych narzędziach umysłu ucieleśnionego oraz ucieleśnionej matematyki, należy wskazać na mechanizmy neuronalne, które mogłyby te funkcje realizować. Teorie Lakoffa i jego epigonów nie mówią o neuronalnym podłożu i konkretnych procesach biofizycznych odpowiedzialnych za zdolność do posługiwania się językiem. Nie podają także sposobu ich obliczeniowego modelowania, który umożliwiłby, choćby w dalekiej przyszłości, przeniesienie języka naturalnego na sztuczny intelekt, tak aby mógł on rozumieć inne osoby żywe lub sztuczne. Postęp w badaniach procesów biofizycznych leżących u podłoża procesów neuronalnych umożliwia zarysowanie procesu tworzenia języka.

Poszukiwanie biologicznego podłoża naszych zdolności językowych wymaga odwołania się do architektury świadomości umysłów naturalnych, ponieważ bez świadomości nie jest możliwe poznanie i rozumienie, a bez nich komunikacja za pomocą języka nie może istnieć. Model architektury zdolnej do rozumienia języka naturalnego obejmuje specyficzną semihierarchiczną strukturę neuronowych pól modelujących postulowanych przez Perlovsky'ego [2001, 2007], w których zachodzą procesy tworzenia reprezentacji neuronalnych postrzeganych obiektów. Na poziomie biologicznym tym neuronowym polem modelującym odpowiadają skupiska wysepek synaptycznych pełniących funkcję rozciągniętych przestrzennie sond ładunkowych zdolnych do reagowania na rozkłady ładunków elektrycznych przynoszonych przez sygnały pobudzające zgodnie z zasadami neuro-elektro-dynamiki Doriana Aura i Mandara S. Joga [Aur, Jog, 2010, 2007]. Oddziaływaniom czasoprzestrzennych „impulsów kierunkowych” (w kategoriach neuro-elektro-dynamiki), generowanych przez pola elektryczne sygnałów i sprzężone kompleksy neuroprzebieżników, towarzyszą odkształcenia struktury łańcuchów białkowych tworzących kanały synaptyczne i tory przepływu ładunków w dendrytach, aksonach i perikarionie neuronu. Zmiany struktury białek wymagają komplementarności rozkładów czasoprzestrzennych impulsów kierunkowych z lokalną topologią mikropól elektromagnetycznych molekuł białkowych. Ta komplementarność jest podstawą znajdowania podobieństw nowo przychodzących pobudzeń sensorycznych z ukształtowaną w toku uczenia strukturą molekularną elementów synaps i neuronów, stanowiącą pamięć krótko-, średnio- i długotrwałą.

Natomiast odkształcenia tej struktury wskutek silnych lub powtarzających się impulsów pobudzających są podstawą plastyczności pamięci umożliwiającą uczenie.

Model uwzględniający te oddziaływania został przedstawiony w eseju *Architektura świadomości* [Galus, 2015a, 2015b, 2015c]. Według tego modelu neuronowe pola modelujące tworzą wielowarstwową strukturę semihierarchiczną, w której pobudzenie rozpoznanych dzięki podobieństwu do zagnieżdżonych w pamięci wzorców przekazywane jest do wyższych warstw (sygnały *bottom-up*). Ponieważ pobudzenie kilkadziesiąt do kilku tysięcy pól synaptycznych potrzebne jest do pobudzenia neuronu wyższej warstwy, to w tym transporcie następuje olbrzymia kompresja informacji, a hierarchia pobudzeń tworzy tzw. impresjony, czyli drzewa pobudzeń odwrócone „koroną” w dół. Nazwa „semihierarchiczna” wiąże się z tym, że tak powstające struktury mają odwrotną hierarchię zwrotnego i równoległego przekazywania pobudzeń z góry na dół (*top-down*). Nazwa „impresjon”, odpowiadająca nazwie „semblion” wprowadzonej przez Kunjumona Vadakkana [2011, 2012a, 2012b], oznacza struktury neuronalne tworzące podłoże biologiczne dla reprezentacji mentalnych elementarnych postrzeżeń na poziomie sensorycznym, aż do pojęć, konceptów, idei i modeli, na wzrastających poziomach struktur pojęciowych definiowanych symbolicznie. Tak więc impresjony, jako neuronalne reprezentacje tego, co postrzegamy lub co zapamiętaliśmy z wcześniejszych doświadczeń, to specyficzne konfiguracje hierarchicznie zorganizowanych struktur „neuronowych pól modelujących” według nomenklatury Perlovsky’ego. Ich organizacja następuje poprzez utrwalenie ścieżek pobudzeń neuronalnych przesyłanych do pól synaptycznych neuronów poprzez kanały synaptyczne, dendryty i aksony, a także wędrujących poprzez somę neuronów. Mamy tu do czynienia ze ścisłą równoważnością impresjonu w sferze biofizycznej z jego reprezentacją mentalną, w sferze psychicznej. Pobudzenie impresjonu oznacza wzbudzenie w naszym umyśle odpowiadającej mu myśli. Bliższy opis roli impresjonów w procesach kognitywnych oraz procesów biofizycznych je konstytuujących można znaleźć w przytoczonych wyżej pracach [Galus, 2015a, 2015b, 2015c, 2018] i literaturze tam cytowanej.

Istotą modelu architektury świadomej są sprzężenia efaptyczne, poprzeczne, pomiędzy impresjonami poprzez sąsiadujące pola synaptyczne. Umożliwia to nie tylko rozpoznawanie podobieństw konfiguracji pobudzeń synaptycznych, ale także asocjacje grup impresjonów oraz iluzje pobudzeń sensorycznych przy okazji rzeczywistego pobudzania impresjonu dotyczącego innej modalności. Impresjony niskich poziomów odpowiadają zapamiętanym prostym wrażeniom zmysłowym, a poziomy wyższe tworzą bardziej rozbudowane struktury i reprezentują złożone obiekty abstrakcyjne. Dopiero na tym wyższym poziomie hierarchii uogólnień można mówić o amodalności reprezentacji mentalnych. Wielkie, zasocjowane sieci impresjonów mogą tworzyć modele sceny lub złożonych procesów. Procedura transponowania sekwencji czasowych zdarzeń na strukturę przestrzenną umożliwia zapamiętywanie epizodów w pamięci epizodycznej, a także ich odtwarzanie. W cytowanej pracy wykazano, że postulowana struktura sieci neuronowej jest zdolna do samouczenia motywowanego. Wskazano na główne motywacje do działania osobników wyposażonych w mózg posiadający postulowaną architekturę. Procedurę uczenia motywowanego, gdzie osobnik działający autonomicznie zdolny jest do samodzielnego formułowa-

nia celów działania, planowania i podejmowania decyzji, przedstawił Janusz Starzyk i współpracownicy. Zaproponował on mechanizm tworzenia „hierarchii bólów”, od prymitywnych bólów zmysłowych aż po „ból psychiczny” o charakterze abstrakcyjnym (unikanie bólu jest uznawane jako silna motywacja do działania). Odpowiadać im musi komplementarna struktura impresjonów na różnych poziomach wyabstrahowania, od poziomu sensorycznego aż do poziomu najbardziej abstrakcyjnych odczuć [Starzyk, 2011; Starzyk i in., 2012; Graham, Starzyk i Jachyra, 2015]. Na poziomie sensorycznym i na niskich stopniach hierarchicznej struktury impresjonów odpowiadają one bezpośrednim doznaniom zmysłowym, postrzeżeniom najprostszych cech obiektów, perceptom pojedynczych obiektów, a na wyższych poziomach stopniowego uogólniania – ideom, złożonym konceptom, modelom i wreszcie ogólnemu modelowi postrzeganego świata, nazywanego światopoglądem. Proces tworzenia impresjonów został opisany bliżej w części 4 niniejszego artykułu.

W przywołanej wyżej pracy *Architektura świadomości* przedstawiono pogląd, że unikanie bólu i zabieganie o doznania przyjemne motywuje do zachowań inteligentnych. Natomiast dla zachowań świadomych i zmierzających do poszerzenia świadomości konieczne jest przejawianie „instynktu rozumienia” bazującego na ciekawości tożsamej z zachowaniami eksploracyjnymi. W części III wymienionej wyżej pracy, w rozdziale 11 (podrozdział *Ciekawość*) oraz w rozdziale 12 (*Kompleksowy model umysłu świadomego*, podrozdział *Zrozumieć procesy rozumienia*) przedstawiono hipotezę, iż podłożem tego instynktu są naturalne właściwości neuronowych pól modelujących. Hipoteza ta opiera się na znanym psychologicznym efekcie odczuwania satysfakcji emocjonalnej, kiedy umysł poznaje coś nowego. Mogą to być nowe wrażenia bliskie bezpośrednim odczuciom zmysłowym, nowe obiekty, na wyższym poziomie generalizacji, lub nowe idee bądź wrażenia estetyczne na najwyższym poziomie naszego światopoglądu czy gustów artystycznych. To uczucie przyjemności z poznawania i rozumienia nowego jest podstawą procesów uczenia, poszukiwań badawczych, eksperymentów technicznych, tworzenia sztuki i rozwoju kultury. Dostrzegli to już starożytni, co znalazło odzwierciedlenie w interpretacji stoickiej eudajmonii lub choćby w opisie jej roli w pismach Arystotelesa.

Poznawanie nowego i włączanie go w konstrukcję wiedzy o świecie i zbiór wcześniejszych doświadczeń nazwano tam procesem rozumienia, a fakt dążenia do eksploracji otoczenia i dostępnych źródeł wiedzy określono jako instynkt rozumienia. Podobnego instynktu ciekawości i poszukiwania zrozumienia doszukuje się na przykład Michał Heller w swej książce *Filozofia nauki* [Heller, 2016]. To dążenie (popęd) napędzane jest odczuwaną satysfakcją emocjonalną, przyjemnością z czystego aktu poznawania. Można kwestionować sens wyodrębnienia tego instynktu, ponieważ kognitywistyka przyjmuje za pewnik, że najsilniejszą motywacją działania naturalnych i sztucznych systemów inteligentnych jest unikanie bólu i poszukiwanie przyjemności. Jednakże eudajmonia zdecydowanie różni się od pospolitej hedonii, czyli bezpośredniego odczuwania przyjemności dzięki naszym zmysłom. O ile bezpośrednia przyjemność zmysłowa dostępna jest niemal wszystkim stworzeniom inteligentnym zaopatrzoną w zmysły, o tyle przyjemność ze zdobywania nowej wiedzy dostępna jest umysłom na wyższym poziomie, zdolnym do budowy złożonego modelu otoczenia poprzez uczenie się i zapamiętywanie przeszłych doświadczeń. Odczuwanie

tego rodzaju przyjemności nie jest związane z podrażnieniem jakichkolwiek zmysłów. Uczucie to generowane jest wewnątrz mózgu wskutek procesów neuronalnych zachodzących w trakcie procesów poznawczych. Można antycypować, iż jest to część przedstawionej wyżej hipotezy, że naturalne sieci neuronowe wykazują dążenie do komplementarności stanów pobudzeń neuronowych i utrwalonych w proteinach wzorców zagnieżdżonych w pamięci. A także, że osiągnięcie stanu komplementarności sygnalizowane jest biochemicznie jako przyjemność napędzająca to dążenie. Byłaby to biofizyczna podstawa instynktu „rozumienia”.

Neuronowe pola modelujące (NPM) osiągają wskazany wyżej cel doprowadzenia do komplementarności stanów poprzez kierowanie i kojarzenie właściwych sygnałów pobudzających z odpowiadającymi im polami synaptycznymi, a także poprzez ciągłe asocjacje w celu maksymalizacji sygnału zgodności wzorców. Sukces tych asocjacji uwalnia mechanizm nagrody, prawdopodobnie hormonalny, co rejestrowane jest w mózgowych ośrodkach przyjemności jako odczuwanie przyjemności w wyniku postrzegania nowości i jako przyjemne poczucie „zrozumienia”. Obiekty do obserwacji lub do rozważań umysłowych, odwzorowane w impresjonach, selekcyonowane są poprzez wybór najintensywniejszych pobudzeń, a następnie kierowane ścieżką skojarzeń pobudzających efektywnie (poprzecznie) lub zstępująco (zwrotnie) do pamięci roboczej, gdzie są uświadamiane. Jak napisano w eseju *Architektura świadomości*, intensywność pobudzeń wynika ze zmian intensywności pobudzeń pól synaptycznych związanych z ruchem i elementami nowości w sygnałach doprowadzanych do NPM. Opisany tam mechanizm selekcji zawartości pamięci roboczej, a przez to naszej świadomości, nazwano mechanizmem „uwagi”. Błądzenie myśli i przeskakiwanie z tematu na temat odbywa się – według Starzyka – dzięki sakadom mentalnym na wzór sakad wzrokowych, kierujących nasz wzrok na istotne obiekty sceny przed naszymi oczyma [Starzyk i Graham, 2015]. Procesy uczenia motywowanego możliwe są tylko w przypadku agenta autonomicznego działającego w dynamicznie zmiennym otoczeniu i zamykają się w kręgu percepcji (sensorycznej) – przetwarzania informacji w celu planowania działania w środowisku (w umyśle świadomym) – wykonania zaplanowanych działań (dzięki efektorom) – powtórnej obserwacji skutków działania. Taki system, uczący się dzięki manipulacjom w środowisku, nazywamy „inteligencją ucieleśnioną”. W przedstawionym procesie cyklicznym budowany jest model otoczenia, świata, w którym inteligentny system się porusza, co jest równoznaczne z uzyskiwaniem wyższego stopnia świadomości. Już nie świadomości czysto percepcyjnej, ale świadomości czasoprzestrzeni, wyodrębnienia w niej własnej osoby, czyli uzyskania samoświadomości. Pierwotna idea Lakoffa, poznania dzięki prostym schematom wyobrażeniowym, ulega znaczącemu rozszerzeniu.

Model umysłu oparty na heurystyce neuronowych pól modelujących zdolnych do tworzenia i kolektywnych współoddziaływań impresjonów nie jest sprzeczny z wcześniejszymi koncepcjami kreowania języka i procesu myślenia z użyciem języka symbolicznego (w odróżnieniu od procesów neuronalnych na poziomie odruchów, prostych odczuć i reakcji na pobudzenia zmysłowe) przez umysł ucieleśniony, począwszy od słynnych K-Lines Marvina Minsky’ego [1980] aż po „ucieleśnione znaczenie” w neuronowej teorii języka Jerome’a Feldmana i Srinivasa Narayanana [2004]. Wykorzystajmy ostatni przykład do wykazania siły prezentowanego modelu. Autorzy wskazują, że

kora przedruchowa (pole 6 Brodmanna) i ciemieniowa u człowiekowatych są sprzężone nie tylko w celu skoordynowanego sterowania motorycznego (uszkodzenie pola 6 powoduje niezdolność ruchów – ataksję), ale także tworzą reprezentacje mentalne motorycznego działania, co pozwala na jego planowanie. Autorzy stawiają hipotezę, że ten zespół ośrodków mózgowych służy jako neuronowe podłoże znaczenia przywiązanego do słów opisujących planowane lub wykonane działanie. Odwołują się do pracy Pulvermuellera [2001], gdzie przytacza on dowody istnienia obwodów sensoryczno-motorycznych służących temu celowi, to jest ugruntowaniu cielesnego znaczenia słów (ucieleśnione znaczenie). Należy jednak wykazać, jak poszczególne znaczenia słów odpowiadające konkretnym działaniom mogą być łączone z tymi ucieleśnionymi w wymienionych strukturach mózgowych pojęciami, które do tych działań się odnoszą. Potrzebna jest teoria, jak wyrażenia łączone są z ich znaczeniem. Feldman i Narayanan opisują prosty eksperyment kojarzenia koincydencyjnego wypowiedzianych jako pojedyncze słowo etykiet (nazw) z odpowiadającym im ruchem i interpretują wyniki na gruncie gramatyki konstrukcyjnej (*construction grammar*). Zastosowany sposób interpretacji wyników ograniczający się do wskazania wielkoskalowych struktur mózgowych niewiele wnosi do zrozumienia sposobu tworzenia skojarzeń znaczeniowych i asocjacji werbalno-motorycznych. Jednakże przytoczone wyniki można wspaniale wykorzystać do uzasadnienia mechanizmu tworzenia impresjonów. Wyniki te wskazują, że impresjony mogą obejmować struktury motoryczne, oraz dowodzą, że asocjacje multimodalne impresjonów rzeczywiście zachodzą. Możliwe jest głębokie objaśnienie otrzymanych wyników, sięgające poziomu procesów biofizycznych w neuronowych polach modulujących. Jak dalej pokażemy, pozwala to poważnie myśleć o zbudowaniu sztucznego agenta autonomicznego, zdolnego do werbalnego meldowania o zamierzonych działaniach i opisu czynności już wykonanych.

Czy architektura impresjonów jest w stanie odkryć strukturę o złożoności większej niż prosty proces Markowa? To trudne, gdyż należałoby pokazać, dlaczego ta hipoteza nie zostaje automatycznie obalona przez lemat o pompowaniu. Odpowiedź na to pytanie wymaga odwołania się do procesu zapamiętywania i przetwarzania ciągów zdarzeń w pamięci epizodycznej. Następuje przy tym konwersja ciągów czasowych pobudzeń neuronowych w przestrzenne rozkłady zmian konformacji protein utrwalających te pobudzenia w polach synaptycznych. Powstają w ten sposób impresjony dynamicznie reprezentujące ciągi zdarzeń, którymi mogą być także fonemy i słowa języka. Sposób powstawania takich impresjonów omówiono dokładniej w pracy Galusa [2015b, s. 255]. Rozpoznawanie struktury napływających kolejno sygnałów przetransponowanych do formy przestrzennej w pamięci epizodycznej następuje poprzez ich porównanie i wykrycie podobieństwa do istniejących w mózgu, a zatem i w umyśle, wzorców powstałych w toku uczenia się języka. To rozpoznawanie ma więc charakter identyczny jak rozpoznawanie wzorców obiektów przestrzennych. Operacja wykrywania podobieństwa, a nie ścisłej struktury sekwencji słów, ogranicza formalną regularność języka.

Cały dorobek eksperymentalny badaczy, na który powołują się Feldman i Narayanan w omawianej wyżej pracy, jest silnym argumentem za występowaniem postulowanych mechanizmów i organizacji pamięci w mózgach naturalnych. Natomiast osadze-

nie impresjonów w procesach biofizycznych neuro-elektro-dynamiki daje znacznie głębsze podstawy do analizy możliwości i ograniczeń ich wzajemnych oddziaływań, a w konsekwencji stwarza szansę dalszych badań nad heurystyką neuronowych pól modelujących, prowadzących do ewentualnej falsyfikacji postulowanej teorii umysłu świadomego. Prace Giacomo Rizzolattiego i współpracowników, cytowane przez Feldmana i Narayanana w przywołanym powyżej artykule, mogą być traktowane jako potwierdzenie rzeczywistego występowania impresjonów w ludzkim mózgu [Rizzolatti i in., 1996 i inne]. Przewagą hipotezy impresjonów w stosunku do wcześniejszych fragmentarycznych teorii jest jej objaśnienie i uzasadnienie fundamentalnymi procesami biofizycznymi, neurologicznymi i psychologicznymi na każdym poziomie analizy funkcjonowania mózgu i generowanego przez mózg umysłu.

Należy zwrócić uwagę na wielką zbieżność prezentowanego modelu umysłu świadomego z teorią metafor językoznawstwa kognitywnego. Postulat umysłu ucieleśnionego zyskuje głębokie uzasadnienie ideą inteligencji ucieleśnionej, wprowadzonej przez badaczy klasycznej sztucznej inteligencji. Pojęcia te nie są tożsame, a jednak prezentowany model wskazuje, jak inteligencja ucieleśniona, gromadząc w pamięci doświadczenie poprzez manipulowanie środowiskiem, zdolna jest do tworzenia neuronalnych korelatów efektów tych doświadczeń, tworząc umysł ucieleśniony. Natomiast lingwistyczne metafory odnajdujemy w strukturach impresjonów, które sprzęgane są poprzez neuronowe pola modelujące, prowadząc do rozległych asocjacji odpowiadających złożonym modelom, ideom i innym abstrakcyjnym konstruktom lingwistycznym wyrażonym w języku symbolicznym. Mechanizm tworzenia metafor odpowiada odnajdywaniu podobieństw między impresjonami o podobnej strukturze. Jak więc widzimy, istnieje ścisły, przyczynowy związek między kompleksowym modelem architektury umysłów świadomych a teorią metafor. Eksperymentalne potwierdzenie za pomocą sondy tetrodowej [Aur i Jog, 2010; Jog i Aur, 2009; Jog, Aur, Connolly, 2007] tworzenia struktur białkowych będących biofizycznym podłożem pamięci hierarchicznej, tożsamej z neuronowymi reprezentacjami postrzeżeń i działającej poprzez wyszukiwanie podobieństwa (a nie identyczności) odbieranych konfiguracji pobudzeń w polach synaptycznych, jest znaczącym wsparciem intuicji Lakoffa i Johnsona. Ich językowe – mentalne metafory mają biofizyczne odwzorowanie w podobieństwach topologii ładunków elektrycznych w polach synaptycznych (NPM) do rozkładów ładunków w impulsach sensorycznych.

Nic nie ogranicza rozległości, szczegółowości i precyzji odwzorowania rzeczywistości poprzez asocjowane zespoły impresjonów, co pozwala tworzyć najbardziej wyrafinowane modele matematyczne. Nowoczesne metody obrazowania pracy mózgu potwierdzają, że podczas procesów umysłowych, dotyczących złożonych struktur, pobudzone są niemal wszystkie rejony kory mózgowej. Najczęściej przeciętnie zdolny umysł nie jest w stanie uświadomić sobie całej konstrukcji i kompletnego modelu, z którym się zapoznaje. Natomiast dzięki zdolności kojarzenia i opisanemu w pracy *Architektura świadomości* mechanizmowi przełączenia uwagi możliwe jest „błądzenie” po różnych aspektach modelowanego problemu. Mamy tu więc zarówno schematy wyobrazeniowe, jak i system aspektowy. Aby „ujrzeć” całość problemu, niezbędne bywa wielokrotne przejrzanie całego schematu. Bardzo w tym pomaga notacja symboliczna, która dzieli cały schemat i jego reprezentację mentalną w postaci

skojarzonego zespołu impresjonów na części odpowiadające subimpresjom i oznaczone jednoznaczными symbolami – słowami lub symbolami matematycznymi. Dzięki temu następuje wyabstrahowanie najistotniejszych relacji i schematów procesów przebiegających pomiędzy cząstkowymi, dobrze zdefiniowanymi pojęciami. Skutkuje to kolosalną kompresją informacji pozwalającą na operowanie bardzo złożonymi schematami. Na szczęście nasze umysły posiadają zdolność kojarzenia schematów wyobrażeniowych zaklętych w impresjach z symbolami dźwiękowymi, graficznymi lub gestykulacyjnymi. W ten sposób tworzą one język mówiony i pisany. Odbywa się to poprzez kojarzenie postrzeganych obiektów z napływającymi sygnałami innych modalności. Podstawowym kryterium kojarzenia jest równoczesność (koincydencja). Uczymy się języka poprzez zademonstrowanie obiektu i równoczesne powiedzenie lub usłyszenie jego nazwy, wskazanie odpowiedniej inskrypcji lub gestu (w przypadku np. języka migowego). Kiedy obiekt ma charakter abstrakcyjny, wówczas wystarczy opis jego cech dla przywołania odpowiedniego schematu wyobrażeniowego i przedstawienie odpowiadającego mu symbolu.

Dzięki istnieniu impresjonów dynamicznych (zmiennych w czasie, lecz zapisanych przestrzennie w strukturze neuronowych pól modelujących, co opisano w cytowanej pracy *Architektura świadomości*) możemy posiadać schematy wyobrażeniowe procesów zapisywane w pamięci epizodycznej. Pozwala to kojarzyć typowe operacje i stosować je w nowych dziedzinach. Mamy wówczas do czynienia nie tyle z metaforami, ile z analogiami. Zresztą funkcja wyszukiwania podobieństwa konfiguracji pobudzeń neuronalnych potrafi obsługiwać zarówno postrzegane metafory, jak i odkrywane analogie. Dzięki temu spostrzeżenie Lakoffa i Núñeza, że: „(...) metafora pojęciowa jest neuronalnie ucieleśnionym, podstawowym mechanizmem poznawczym, który pozwala na użycie inferencyjnej struktury jednej dziedziny do wnioskowania o innej dziedzinie” znajduje ścisłe neurobiologiczne uzasadnienie. Słuszna jest zatem także ich intuicja, iż pojęcia i idee, które powstają w procesie metaforyzacji, nie stanowią wyłącznie obiektów matematycznych, ale mogą być również podstawą systemu matematycznego wnioskowania. Schematy wyobrażeniowe procesów, stosując terminologię Lakoffa, też mogą uzyskiwać nazwy – symbole, które możemy stosować metaforycznie lub tworząc bardziej złożone konstrukty stosowane w podobny sposób w wielu dziedzinach. Z punktu widzenia roli impresjonów można stwierdzić, że podstawą poznania jest raczej proces wyszukiwania analogii i podobieństw, i właśnie w ten sposób proponuję poszerzyć pojmowanie metafor Lakoffa.

Należy także podkreślić zgodność hierarchicznego modelu metafor z hierarchiczną strukturą impresjonów. Zarówno utrwalone proste, bezpośrednio zmysłowe wrażenia i pojęcia metaforyczne na poziomie *qualiów*, jak i odpowiadające im impresjony zagnieżdżone są w najniższych warstwach pól neuronowych sprzężonych bezpośrednio z naszymi zmysłami. Powoduje to głębokie osadzenie tworzonych na wyższych poziomach pojęć abstrakcyjnych w tych bezpośrednich, subiektywnych wrażeniach zmysłowych (*symbol grounding*). Umożliwia to proporcjonalnie głębokie, wieloaspektowe rozumienie abstrahowanych pojęć. Sztuczne systemy inteligentne, nawet wyposażone w potężne relacyjne bazy danych, w istocie nie rozumieją pojęć, którymi operują według narzuconych przez programistów reguł, ponieważ nie wbudowują tych pojęć w szerszy model świata oraz nie odnoszą tych pojęć do elementarnych wrażeń

zmysłowych, jakie kontakt z obiektami symbolizowanymi przez te pojęcia wywołuje. Nie posiadają naszych *qualiów*, nie posiadają światopoglądu. Przeto odmawiamy im świadomości, odmawiamy im celowego działania i wiemy, że nawet jeśli rozumiemy język, którym do nas przemawiają, to one same swego języka nie rozumieją. Inne osiągnięcia modelu wyjaśniające takie aspekty jak podświadomość, intuicja, wolna wola omówiono szerzej w książce *Świadomość? Ależ to bardzo proste!* [Galus, Starzyk 2018].

Gdzie w prezentowanym modelu miejsce dla neuronów lustrzanych i roli imitacji, tak istotnych dla obserwacji kumulatywnego charakteru tworzenia kultury w interakcjach społecznych? Model kognitywnej architektury samoświadomej wynikającej ze specyficznego ustrukturyzowania neuronowych pól modelujących silnie wspiera hipotezę neuronów koncepcyjnych (*concept cells*, *grandmother cells*). Można przyjąć w pewnym uproszczeniu, że tzw. neurony lustrzane to właśnie neurony koncepcyjne, które pobudzone sensorycznie indukują wrażenia i reakcje podobne, jakie odczuwa obserwowany sąsiedni osobnik. Nie dowodzi to istnienia neuronów lustrzanych w sensie entuzjastycznych doniesień z początków lat 90. ubiegłego wieku dokonanych przez zespół Giacomo Rizzolattiego z Parmy [Di Pellegrino i in., 1992; Gallese i in., 1996; Rizzolatti i in., 1996]. Nie podważa zastrzeżeń zgłaszanych przez Hickoka [2009], Pascolo [Pascolo, Ragogna i Rossi, 2009] i Dinsteina [Dinstein i in., 2008]. Istnienie specyficznych neuronów lustrzanych o wyspecjalizowanych funkcjach nie zostało udowodnione. Nikt jednak nie podważył wyników badań behawioralnych prowadzonych w zespole Rizzolattiego przez ponad 20 lat. Wątpliwości może budzić jedynie ich interpretacja. Przedstawione alternatywne wyjaśnienie obserwowanych zachowań najpierw makaków, a potem i ludzi poprzez występowanie hierarchicznych struktur nazywanych impresjonami wydaje się dużo lepiej uzasadnione. Grupy komórek na szczytach hierarchii impresjonu możemy nazywać neuronami koncepcyjnymi, a reakcje całych impresjonów zewnętrznie pobudzanych poprzez doznania zmysłowe możemy identyfikować z reakcjami „neuronów lustrzanych”, które w tym przypadku nie będą już pojedynczymi specyficznymi neuronami, a złożonymi, lecz typowymi strukturami naszej sieci neuronowej. Zdając sobie sprawę z trwających kontrowersji wokół koncepcji neuronów koncepcyjnych, nie możemy ignorować eksperymentalnych potwierdzeń występowania tego typu struktur w mózgach ludzkich przedstawionych na przykład przez Quiroga i współpracowników [2013, 2005]. Istnienie impresjonów jest potwierdzone wykryciem struktur hierarchicznych w całym obszarze kory mózgowej. Taką budowę ma kora wzrokowa, w której wyróżniamy warstwy od V1, V2 poprzez V4... V7, gdzie można wyodrębnić jeszcze mapy retinotopowe, aż po ponad 20 kolejnych warstw pobudzanych przez bodźce wzrokowe. Podobnie jest dla innych modalności, w tym dla zmysłu słuchu: od warstw kory pierwszorzędowej o strukturze tonotopowej, poprzez korę słuchową drugo- i trzeciorzędową, aż do obszarów F4 kory przedczołowej i czołowej pobudzanej dźwiękowo, lecz jednocześnie asocjującej bodźce innych modalności.

W przywoływanej już pracy *Architektura świadomości* można znaleźć uzasadnienie, dlaczego inteligentne osobniki uczące się w podobnym środowisku będą posiadały podobne *qualia*, zwłaszcza jeśli dysponować będą podobnymi zmysłami. Analogicznie argumentować można, że przy podobnym poziomie świadomości i doświadcze-

nia (w wyniku procedur uczenia) dysponowały będą podobnym systemem pojęciowym. Być może ich reakcje motoryczne będą różne, lecz schematy wyobrazeniowe, rozumienie reakcji i przewidywanie następnych działań powinny być adekwatne do doświadczanego otoczenia fizycznego i społecznego. Odpowiadające im reprezentacje mentalne w postaci impresjonów także powinny mieć podobną strukturę i z tego powodu, przez analogię do neuronów lustrzanych, można je nazywać impresjonami lustrzanymi.

Zdolność do imitacji nie wynika wprost z przytaczanego modelu świadomości. Warto zwrócić uwagę, że uczenie społeczne przez imitację w mniejszym stopniu przyczynia się do rozwoju języka, a w większym do jego upowszechniania. Mimo to olbrzymie znaczenie zdolności do imitacji nie powinno być kwestionowane. Zdolność ta jest szeroko wykorzystywana w królestwie zwierząt. Bezdyskusyjne jest jej znaczenie ewolucyjne i także zapewne jest jej pochodzenie. Można ją wiązać z instynktem ciekawości i eksploracji. Jak pokazano w wymienionej pracy, instynkt eksploracji powściągnięty jest strachem przed dyskomfortem, bólem, śmiercią. Naśladowanie innych osobników może znacznie zwiększać odwagę podejmowania nowych działań. Podejmowane ryzyko często się opłaca, bo sprzyja eksploracji nowych terenów i nowych zasobów. Zdolność imitacji tworzy zatem przewagę ewolucyjną i powinna podlegać pozytywnej selekcji. Bardzo pomocne do tego celu mogą być impresjony lustrzane. Trafna jest ocena Brożka i Hohola, że mistrzostwo ludzi w imitowaniu zachowań osobników własnego i innych gatunków jest jedną z podstaw sukcesu naszej kultury. Inne to zdolność do kumulatywnego rozwoju kultury dzięki tworzonym, a następnie także utrwalanym symbolom [Brożek i Hohol, 2014]. To właśnie impresjony utrwalające pojęcia i reguły języka można uznać za poszukiwany przez Chomsky'ego „moduł językowy”.

Niezwykle podobieństwo sieci neuronowej w świecie zwierzęcym nasuwa pytanie: dlaczego inne zwierzęta nie tworzą symbolicznych języków komunikacji? Należymy przecież do tego samego świata biologicznego. Wszak wiele badań nad zachowaniem neuronów prowadzi się na neuronach ślimaków. Mózgi ssaków wykazują wielkie podobieństwo morfologiczne, nie mówiąc już o mózgach makaków czy małp człekokształtnych. Można przypuszczać, że w ich mózgach powstają złożone impresjony reprezentujące *qualia* podobne do naszych. Ich załączki systemu komunikacyjnego wskazują na zdolność do tworzenia pojęć-symboli o pewnym stopniu abstrakcji. W czym leży różnica powodująca tak dramatyczne skutki?

Moim zdaniem można przypuszczać, że różnica kryje się w liczbie poziomów, na których integrowana jest struktura impresjonów, i w rozległości asocjacji, które mogą pomiędzy nimi zachodzić. Przesłanki przedstawiane w przywoływanej pracy Galusa [2015a] wskazują, że sieć neuronowa wymaga nie dwóch lub trzech poziomów przetwarzania, ale kilkudziesięciu. Być może nawet do 20–40 poziomów. Każdy dodatkowy poziom gigantycznie powiększa liczbę kombinacji i permutacji sposobów powiązania neuronowych pól modelujących tworzących struktury impresjonów. Być może w procesie ewolucyjnym nawet niewielka mutacja powiększająca powierzchnie kory mózgowej umożliwiła jednocześnie kierowanie pobudzeń neuronowych do kolejnych warstw i pól modelujących, powodując powstawanie jeszcze bardziej wyabstrahowanych pojęć – konceptów – modeli. Nie mniejszą rolę w zdolności do

abstrakcyjnego myślenia i posługiwania się językiem odgrywa zdolność do asocjacji i kojarzenia pojęć. Właściwości te są zależne od skuteczności sprzężeń efaptycznych na poziomie pól synaptycznych, od zasięgu połączeń aksonowych i dendrytycznych pomiędzy współpracującymi warstwami kory i pól neuronowych, od dostępności molekuł właściwych białek w kanałach synaptycznych, dendrytach i aksonach oraz od subtelności funkcjonowania, a szczególnie od poziomów wrażliwości ustanawianych przez mechanizm „uwagi”, czyli progu dyskryminacji pobudzeń pól synaptycznych. Czynniki te decydują o osobniczych zdolnościach do kojarzenia, asocjowania i zapamiętywania wiedzy i mogą się zmieniać wraz ze wzrostem, rozwojem, nabywaniem doświadczeń, starzeniem oraz pod wpływem zmian patologicznych, diety i używek. Działanie większości tych czynników jest neurologicznie znane. Na temat zróżnicowania i pożądanej struktury białek tworzących strukturę synaptyczno-neuronową niewiele możemy powiedzieć, poza tym, że u ludzi zdolności do zapamiętywania znacząco się różnią. Także w niewielkim stopniu przebadano przebieg potężnych wiązek aksonów łączących poszczególne pola neuronowe i kierujących pobudzenia neuronowe do wyspecjalizowanych struktur mózgu. Część tych wiązek kieruje zwrotnie sygnały z wysokich do niższych pól i to nie zawsze w ramach tej samej modalności. O ich zasięgu możemy wnioskować na podstawie obrazów dostarczanych przez nowoczesne metody obrazowania aktywności mózgu. Także i one odznaczają się dużą zmiennością osobniczą. Natomiast różnice międzygatunkowe nie zostały objaśnione w stopniu pozwalającym nam na wyciąganie ogólnych wniosków.

Możemy stwierdzić, iż poszczególne osobniki różnią się w istotnym stopniu indywidualnymi zdolnościami. Mózg ludzki ze względu na swoją złożoność i powierzchnię kory wydaje się doskonałym podłożem dla wielopoziomowych struktur impresjonów tworzących złożony system pojęciowy.

4. Język sztucznych sieci neuronowych

Czy możliwe jest, aby sztuczne systemy inteligentne opanowały naturalny, złożony język symboliczny? Czy mogłoby to nastąpić na drodze samouczenia lub choćby poprzez imitację naszego ludzkiego języka? Czy mogłyby one posługiwać się nim ze zrozumieniem? Czy będą wówczas posiadały samoświadomość?

Dotychczas budowane systemy sztuczne rozpoznawania i syntezy mowy nie spełniają tych warunków. Wysoce wyspecjalizowane algorytmy znanych chatbotów pozwalają na rozpoznawanie znacznego zasobu słów określonego języka, zaś inne algorytmy dobierają odpowiedzi, a bardzo często także pytania imitujące konwersację w języku naturalnym. Zintegrowane z chatbotem systemy eksperckie pozwalają nawet na uzyskanie w wyniku takiej konwersacji użytecznych informacji. Jednak zdajemy sobie sprawę, że to tylko imitacja rozmowy, ponieważ chatbot, z którym konwersujemy, nie rozumie swych własnych wypowiedzi. Nie uczy się, poza wyznaczoną dziedziną. Nie posiada kreatywności pozwalającej udzielić odpowiedzi wykraczającej poza deterministycznie zaprogramowany repertuar reakcji na stawiane mu pytania. Często antropomorfizacja chatbota rozumiana jest jako wprowadzenie kontrolowanych elementów przypadkowych o charakterze losowym lub chaotycznym, aby zamaskować

algorytmicznie zdeterminowany charakter jego odpowiedzi. Trudności, jakie piętrzą się przed projektantami systemów chatbotowych, znacznie przekraczają techniczne trudności związane z szybkością przetwarzania informacji, pojemnością wymaganej pamięci, dostępem do zasobów wiedzy, czy też zdolnością przetwarzania olbrzymich zbiorów danych. Podstawowa trudność kryje się w tym, że rozumienie języka wymaga ugruntowania pojęć, które są konieczne do formułowania odpowiedzi, w szerokim, wielowarstwowym kontekście wiedzy ogólnej, to zaś wymaga świadomości. Niezbędne jest posiadanie modelu otaczającej rzeczywistości, a niekiedy modelu całego świata, który nas otacza. Potrzebna jest zdolność do uczenia się, a jako skutek uczenia – zdolność do stałego poszerzania i pogłębiania tego modelu. Zdolność do ciągłej rozbudowy modelu także w wyniku pozyskiwania nowej wiedzy w trakcie prowadzonej konwersacji. Dotychczas nie pojawiały się koncepcje, jak mógłby powstać algorytm dający systemowi świadomość i zdolność rozumnej rozmowy.

Do realizacji tego celu nie jest konieczne odtworzenie wszystkich połączeń i struktury ludzkiego mózgu, tak jak do naśladowania ptasiego czy owadziego latania nie jest konieczne odtworzenie błoniastych lub pokrytych piórami skrzydeł. Natomiast celowe jest zrozumienie architektury sieci neuronowej zdolnej do generacji świadomości. Wiele wskazuje na to, że hierarchiczna struktura impresjonów najskuteczniej może powstawać w wielowarstwowym sieciach neuronowych. Potrzebna nam jest architektura neuronowych pól modulujących i znajomość heurystyki, według której pobudzenia z najniższych warstw sensorycznych przekazywane są do wyższych warstw, gdzie tworzone są pojęcia, koncepty i modele, jak to opisano w cytowanej powyżej pracy *Architektura świadomości* [Galus, 2015a, 2015b, 2015c]. Coraz szersza grupa kognitywistów wyraża wątpliwość, czy skuteczne asocjowanie i porównywanie współoddziałujących równoległe, analogowo i nieliniowo, wielowarstwowym struktur przetwarzania i zapamiętywania może być emulowane lub symulowane przez przetwarzanie szeregowo dyskretnych, a tym bardziej binarnych sygnałów w komputerach sekwencyjnych. W każdym razie digitalizacja licznie, równocześnie przetwarzanych sygnałów, o wielkiej dynamice i szybkozmiennej charakterystyce czasowej, nieodmiennie stawia przed konstruktorami problem potęgowej złożoności kombinatorycznej procesu ich rozpoznawania i przetwarzania [Perlovsky, 1998]. Stąd łatwe usprawiedliwienie dotychczasowych niepowodzeń, podczas gdy równocześnie wzbudzane są nadzieje, że nowe generacje superkomputerów o pentaflapowej wydajności pozwolą na rozwiązanie tego problemu.

Na szczęście powstają koncepcje oparcia systemów sztucznych o hierarchiczną, wielowarstwową strukturę koneksjonistyczną, zaprojektowaną do przetwarzania równoległego, gdzie mamy do czynienia jedynie z liniową złożonością kombinatoryczną, którą można pokonać za pomocą dostępnych już obecnie technologii. Taką ideę przedstawił Adrian Horzyk w formie struktury hierarchicznej sieci neuronowej utworzonej przez specyficzne neurony, których wzajemne oddziaływania modelowano grafami umożliwiającymi łatwe programowanie informatycznej struktury odpowiadającej przyjętej koncepcji funkcjonowania takiej sieci. Jak pisze autor: „Różne kombinacje powiązanych asocjacyjnie neuronów umożliwiają definiowanie prostych i skomplikowanych obiektów oraz relacji między nimi” [Horzyk, 2013, s. 52]. Wyróżniono relacje pomiędzy obiektami o charakterze asocjacyjnego podobieństwa, następstwa, kon-

tekstu, definiowania i tłumienia. Kombinacje tych własności charakteryzujące relacje między neuronami tworzą powiązane asocjacyjnie struktury reprezentujące obiekty. Struktury te autor nazywa „semasselami” – jednostkami semantyczno-skojarzeniowymi. Zdolność do ich tworzenia i stabilnego zapamiętywania wynika z własności neuronu, którego model inspirowany biologicznie przedstawiono w omawianej pracy, ale i z własności samego systemu skojarzeniowego. Istotną cechą modelu neuronu zdolnego do funkcjonowania według opisanego systemu asocjacyjnego są zmienne charakterystyki czasowe jego wrażliwości na bodźce. Umożliwia to konkurencyjnym pobudzeniom tworzenie nowych semaseli konkurujących o uwagę. Trudno nie zauważyć zbieżności cech tak modelowanej inteligencji skojarzeniowej z postulowaną architekturą zdolną do wytworzenia świadomości i samoświadomości. Neurony z ich polami presynaptycznymi, wyjściami dendrytycznymi i aksonami odpowiadają neuronowym polom modelującym. Ich podstawowym zadaniem jest wykrywanie podobieństwa, co odpowiada relacjom podobieństwa. Relacje asocjacyjne odpowiadają heurystyce neuronowego pola modelującego regulującej przekazywanie pobudzeń zarówno do warstw wyższych, zgodnie z relacjami definiującymi, jak i do warstw sąsiednich, w których ustalają się relacje kontekstowe. Tworzone w ten sposób semaselle są ścisłym odpowiednikiem impresjonów z ich neuronami koncepcyjnymi na szczycie hierarchii (definiującymi pojęcia-obiekty). Model architektury świadomości przewiduje zarówno relacje następstwa asocjacyjnego, poprzez strukturę pamięci epizodycznej, jak i efektywne relacje tłumienia, dla zachowania dynamiki systemu bez utraty jego stabilności.

Przedstawiona koncepcja modelowania o walorach informatycznej wykonalności nie aspiruje do tworzenia systemu świadomie posługującego się językiem naturalnym. Jednakże model ten wykazał praktyczną skuteczność w realizacji prostych funkcji sortowania i klasyfikatorów. Wskazano, w jaki sposób mogą się tworzyć grafy wiedzy wykorzystywane do automatycznej kategoryzacji i generalizacji. Te procesy są podstawą sztucznej inteligencji skojarzeniowej, do której dążył autor.

Zbieżność tej koncepcji z architekturą mózgu naturalnego stwarza możliwość zbudowania sztucznego systemu zdolnego do świadomego działania, a jego grafowa struktura dostosowana do równoległego przetwarzania informacji wzbudza nadzieję na możliwość szybkiej realizacji technicznej takiego przedsięwzięcia. Tym bardziej, że technologia dostarcza podzespołów doskonale nadających się do roli elektronicznego podłoża do zdefiniowanych w modelu Horzyka funkcji [Benjamin i in., 2014; Merolla i in., 2014]. Zaproponowany model sieci neuronowej posiada dostateczną elastyczność, aby uzupełnić go o funkcje niezbędne do tak sformułowanego zadania. Konieczne uzupełnienie dotyczy wprowadzenia motywacji do działania, uczenia się i planowania akcji według własnego systemu wartości. Wydaje się, że odpowiednim rozwiązaniem będzie zastosowanie algorytmów uczenia motywowanego zdolnością do odczuwania dyskomfortu (ból w terminologii kognitywistycznej) oraz przyjemności, gdy potrzeby systemu będą zaspokajane. Należy także rozważyć wprowadzenie motywacji w postaci zdolności do odczuwania emocji pozytywnej (być może tożsamej z przyjemnością), w przypadku osiągnięcia wysokich wartości wykrywanych podobieństw w rozległych polach synaptycznych i impresjonach, co będzie odpowiadać satysfakcji ze zrozumienia nowych bodźców. Ważnym elementem systemu

świadomego jest ugruntowanie wiedzy symbolicznej w bezpośrednich, subiektywnych wrażeniach zmysłowych, co wymaga tworzenia impresjonów, których hierarchia rozpoczyna się od prostych pobudzeń sensorycznych. Jeśli wybierzemy zestaw sensorów o charakterystykach podobnych do ludzi, to zaistnieją szanse na powstawanie impresjonów podobnych do ludzkich, co znacznie ułatwi przyszłą komunikację z samoświadomymi maszynami. Z pewnością charakterystyki czasowe zmian wrażliwości na bodźce będą musiały być ustalone eksperymentalnie w zależności od dynamiki zmian amplitudowych i czasowych sygnałów pojawiających się w środowisku, w którym miałyby działać system. Niewykluczone, że charakterystyka tych zmian powinna posiadać kilka stałych czasowych odpowiadających nie tylko zmianom rozmiaru perikarionu, ale także habituacji i dyshabituacji neuronu, w przypadku okresowego pojawiania się bardzo silnych lub bardzo słabych sygnałów. Eksperymentalnego doboru wymagać będzie też zasięg połączeń asocjacyjnych i grupowanie połączeń według reguły rzadkich połączeń (*sparse coding*). System musi posiadać możliwość uczenia się w złożonym środowisku poprzez możliwość bezpośredniego manipulowania obiektami w tym środowisku i sprawdzania efektów swego działania. Pozwoli to na wykształcenie odpowiednich impresjonów pamięci proceduralnej umożliwiających skuteczne działanie.

Nasuwa się wątpliwość, czy spełnienie tych założeń nie wymusi stosowania superkomputerów o jeszcze większych mocach niż klasyczna AI. Wydaje się, że takie niebezpieczeństwo nie występuje. W proponowanej architekturze jeden neuron może uczestniczyć w tworzeniu wielu impresjonów. W warunkach rzadkich połączeń do reprezentacji złożonych obiektów wystarczy stosunkowo niewielka liczba neuronów. Nigdzie nie jest powiedziane, że *qualia* sztucznego systemu muszą być hi-fi oraz HD. To pola wzrokowe, słuchowe i dotykowe zajmują dużą część kory mózgowej, a te da się zminimalizować. Ważniejsza jest liczba i rozległość połączeń oraz liczba warstw hierarchii tworzącej impresjony. Jeszcze ważniejsze będą procedury uczenia, podczas których muszą wytworzyć się podstawowe *qualia*, odpowiednie kwantum wiedzy, model rzeczywistości i wreszcie światopogląd. Jeśli elementem tej rzeczywistości będą osoby używające mowy w celu komunikacji, a sztuczny system posiadał będzie efekторы w postaci syntezatora mowy, to świadoma sztuczna inteligencja będzie miała szanse nauczyć się mówić. Jej sprawność komunikacyjna będzie uzależniona od bogactwa wytworzonych impresjonów odzwierciedlających jej wiedzę o świecie i wprawę w posługiwaniu się językiem jako specyficznym narzędziem manipulacji tymi obiektami otoczenia, które zdolne są do komunikacji. Bogaty język raportowania oraz wymiany informacji o swych zamiarach, stanach emocjonalnych i potrzebach ułatwi nam rozpoznanie, że system rzeczywiście posiada świadomość.

Nowo zbudowany system nie będzie tworem gotowym do użytku. Będzie musiał długo zdobywać odpowiednią wiedzę i nabywać wprawę w jej wykorzystaniu, w posługiwaniu się własnym ciałem i umysłem. Musi mieć czas, żeby nauczyć się mówić. Musi mieć czas, żeby tworzyć nowy język matematyki. Miejmy nadzieję, że będzie to trwało krócej niż w przypadku ludzi.

Czy sztuczny system samoświadomy może osiągnąć stopień inteligencji i świadomości taki jak ludzie? Czy proponowany model architektury umysłu naturalnego oraz sztucznego pozwala dodać cokolwiek do dotychczasowych spekulacji? Najistotniejszą

korzyścią z przedstawionych tu modeli jest dostrzeżenie, że świadome posługiwanie się językiem wymaga stosunkowo wysokiego stopnia świadomości. Ludzie mają bardzo różną świadomość i pewnie nastąpi okres, kiedy inteligencja umysłów sztucznych zbliży się do średniego poziomu, jaki obserwujemy u ludzi. Natomiast jest bardzo mało prawdopodobne, żeby zatrzymała się na tym poziomie. Przedstawiony model nie zawiera żadnych ograniczeń w powiększaniu obszarów pamięci neuronowych pól modulujących ani zwiększaniu liczby poziomów przetwarzania. Szybkość i niezawodność układów elektronicznych zapewni większą sprawność systemów sztucznych. Sprawne układy homeostatyczne zapewnią wydajną pracę bez zmęczenia i związanych z tym przerw na jakiś rodzaj snu. Jednakże nie należy spodziewać się cudu dramatycznego przekroczenia ludzkich zdolności intelektualnych. Nawet najbardziej abstrakcyjne pojęcia wymagają osadzenia w *qualiach*, a te, jak wspomnieliśmy wyżej, będą podobne do naszych, jeśli sztuczne systemy będą przebywały w podobnym środowisku. Oczywiście, jeśli wyposażymy je w doskonalsze zmysły – w rentgenowski wzrok, zdolność odczuwania pola magnetycznego czy czegokolwiek jeszcze, to ich bogactwo, różnorodność wrażeń się zwiększy. Może to spowodować powstawanie nowych pojęć, pojawią się nowe słowa, trudno dla nas zrozumiałe. Ale i my dzisiaj jesteśmy zdolni do poszerzenia naszych doświadczeń zmysłowych poprzez stosowanie narzędzi. Nowe postrzeganie świata przez maszyny nie będzie więc dla nas szokiem. Na najwyższych poziomach abstrakcji opisujących modele naszego świata także nie można się spodziewać nagłego oderwania ich poziomu złożoności od modeli przez nas stosowanych. Każda generalizacja w kierunku modelu wyższego poziomu musi opierać się na danych. Zdajemy sobie sprawę, że w świecie fizyki mikro- i makroświata takie dane coraz trudniej nam będzie uzyskiwać. Oczywiście sprawne systemy inteligentne być może pozwolą nam zaplanować nowe eksperymenty, a w ogromnych zbiorach danych dostrzec regularności niedostępne dla naszych umysłów. Nie będzie to jednak proces gwałtowny. Jeszcze długo ludzkie umysły będą nadażały za interpretacjami sprawniejszych umysłów, tak jak uczeń nadaża za swym nauczycielem. Niestety, biologiczna ewolucja mózgu przebiega zbyt wolno i kiedyś jego zdolność adaptacyjna okaże się niewystarczająca. Pozostanie nam wówczas zintegrować nasz mózg ze sztucznym układem wspomagającym. Jeszcze później biologiczny dodatek do supersprawnej maszyny okaże się zupełnie niepotrzebny.

BIBLIOGRAFIA

- Aur, D. i Jog, M.S. (2006). Building spike representation in tetrodes. *Journal of Neuroscience Methods*, 157(2): 364–373.
- Aur, D. i Jog, M.S. (2007). Neuronal spatial learning. *Neural Processing Letters*, 25(1): 31–47.
- Aur, D. i Jog, M.S. (2010). *Neuroelectrodynamics – Understanding the Brain Language*. Amsterdam: IOS Press.
- Benjamin, B.V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A.R., Bussat, J.-M., Alvarez-Icaza, R., Arthur, J.V., Merolla, P.A. i Boahen, K. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, 102(5), May.

- Bonini, C.P. (1963). *Simulation of Information and Decision Systems in the Firm*. Englewood Cliffs, NJ: Prentice-Hall.
- Brożek, B., Hohol, M. (2014). *Umysł matematyczny*. Kraków: Copernicus Center Press.
- Chomsky, N. (1982). *Zagadnienia teorii składni*. Wrocław: Ossolineum.
- Dinstein, I., Thomas, C., Behrmann, M. i Heeger, D.J. (2008). A mirror up to nature. *Current Biology*, 18(1): R13–18.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91: 176–180.
- Gallese, V., Fadiga, L., Fogassi, L. i Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2): 593–609.
- Gallese, V. i Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12).
- Gallese, V. (2009). Ucieleśniona symulacja: Od neuronów po doświadczenie fenomenologiczne. W: A. Klawiter (red.). *Formy aktywności umysłu. Ujęcie kognitywistyczne*, t. 2: *Ewolucja i złożone struktury poznawcze*. Warszawa: Wydawnictwo Naukowe PWN.
- Feldman, J. i Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89(2): 385–392.
- Galus, W.L. (2015a). Architektura świadomości. Część I: Logika i morfologia sieci neuronowej. *Roczniki Filozoficzne*, 63(1): 139–171.
- Galus, W.L. (2015b). Architektura świadomości. Część II: Struktura molekularna i biofizyka pamięci. *Roczniki Filozoficzne*, 63(2): 237–261.
- Galus, W.L. (2015c). Architektura świadomości. Część III: Wola i sens istnienia. *Roczniki Filozoficzne*, 63(3).
- Galus, W.L. i Starzyk, J. (2018). *Świadomość? Ależ to bardzo proste!* Warszawa: BEL Studio.
- Graham, J., Starzyk, J.A. i Jachyra, D. (2015). Opportunistic behavior in motivated learning agents. *IEEE Transactions on Neural Networks and Learning Systems*, 26(8): 1735–1746.
- Heller, M. (2016). *Filozofia nauki*, wyd. III. Kraków: Petrus.
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21(7): 1229–1243.
- Horzyk, A. (2013). *Sztuczne systemy skojarzeniowe i asocjacyjna sztuczna inteligencja*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Jog, M.S., Aur, D. i Connolly, C.I. (2007). Is there a tipping point in neuronal ensembles during learning? *Neuroscience Letters*, 412(1): 39–44.
- Jog, M.S. i Aur, D. (2009). A theoretical information processing-based approach to basal ganglia function. *Advances in Behavioral Biology*, 58, Part 2: 211–222.
- Lakoff, G. (1987). *Women, Fire and Other Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lakoff, G. i Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Book.
- Lakoff, G. i Johnson, M. (2010). *Metafory w naszym życiu*. Warszawa: Aletheia.
- Lakoff, G. i Núñez, R.E. (2000). *Where Mathematics Comes From: How the Embodied Brain Brings Mathematics into Being*. New York: Basic Books.
- Merolla, P., Arthur, J., Alvarez, R., Bussat, J.-M. i Boahen, K. (2014). A multicast tree router for multichip neuromorphic systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61(3).
- Minsky, M. (1980). K-Lines: A theory of memory. *Cognitive Science*, 4: 117–133.
- Pascolo, P.B., Ragogna, R. i Rossi, R. (2009). The mirror-neuron system paradigm and its consistency. *Gait Posture*, 30(Suppl. 1): 65.
- Pawelec, A. (2005). *Znaczenie ucieleśnione. Propozycje kręgu Lakoffa*. Kraków: Universitas.
- Perlovsky, L.I. (1998). Conundrum of combinatorial complexity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6), June.

- Perlovsky, L.I. (2001). *Neural Networks and Intellect: Using Model Based Concepts*. New York: Oxford University Press.
- Perlovsky, L.I. (2007). The knowledge instinct. W: L.I. Perlovsky, R. Kozma (red.). *Neurodynamics of Cognition and Consciousness: Understanding Complex Systems*. New York: Springer.
- Pulvermueller, F. (2001). Brain reflection of words and their meaning. *Trends in Cognitive Sciences*, 5(12).
- Quiroga, R.Q., Fried, I. i Koch, Ch. (2013). Brain cells for grandmother. *Scientific American.com*, 31 February.
- Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C. i Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045): 1102–1107.
- Ramachandran, V.S. (2012). *Neuro nauka o podstawach człowieczeństwa. O czym mówi mózg?* Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Rizzolatti, G. i in. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3: 131–141.
- Starzyk, J.A. (2011). Motivated learning for computational intelligence. W: B. Igelnik (red.). *Computational Modeling and Simulation of Intellect: Current State and Future Perspectives*. Hershey, PA: IGI Publishing.
- Starzyk, J.A., i Graham, J. (2015). MLECOG – Motivated Learning Embodied Cognitive Architecture. *IEEE Systems Journal*, July. Dostępny online: 10.1109/JSYST.2015.2442995.
- Starzyk, J.A., Graham, J.T., Raif, P. i Tan, A.-H. (2012). Motivated learning for autonomous robots development. *Cognitive Science Research*, 14(1): 10–25.
- Vadakkan, K.I. (2011). Processing semblances induced through inter-postsynaptic functional LINKs, presumed biological parallels of K-lines proposed for building artificial intelligence. *Frontiers in Neuroengineering*, 4(8).
- Vadakkan, K.I. (2012a). Framework of consciousness from semblance of activity at functionally LINKed postsynaptic membranes. *Frontiers in Psychology*, 1: 168.
- Vadakkan, K.I. (2012b). The nature of “internal sensations” of higher brain functions may be derived from the design rules for artificial machines that can produce them. *Journal of Biological Engineering*, 6(1): 21.
- Waal, F. de (2014). *Bonobo i ateista. W poszukiwaniu humanizmu wśród naczelnych*. Kraków: Copernicus Center Press.