Jak reprezentowane są pojęcia w mózgu i co z tego wynika.

Włodzisław Duch,
Katedra Informatyki Stosowanej, Uniwersytet Mikołaja Kopernika,
ul. Grudziądzka 5, 87-100 Toruń,
http://www.is.umk.pl/~duch

1. Czym są pojęcia?

A1. Pojęcia zwykle rozumiane są jako mentalne symbole niosące pewne znaczenie, odnoszące się do jakiejś kategorii zdarzeń, klasy rzeczywistych bądź abstrakcyjnych obiektów. Pełne zrozumienie natury pojęć wymaga analizy aparatu poznawczego, który je tworzy i się nimi posługuje. Rozumienia natury pojęć nie można więc oddzielić od procesów poznawczych, a w szczególności myślenia. Z psychologii rozwojowej wiemy (Gopnik, 2004), że podstawowe pojęcia dotyczące świata, takie jak ciągłość istnienia obiektów, które mogą chwilowo znikać z pola widzenia dziecka, rozumiane są dość wcześnie, zanim jeszcze niemowlę potrafi je wyrazić w symbolicznej formie. Dziecko potrafi godzinami chować i wyciągać zabawkę, utrwalając sobie pojęcie ciągłości istnienia. Ontologia podstawowa, konieczna do komunikacji, obejmuje tysiące pojęć, które każdy człowiek zna ze swojego doświadczenia, głęboko zinternalizowanych we wczesnym dzieciństwie.

A2. Autobiografia Hellen Keller, która w 19 miesiącu życia utraciła wzrok i słuch, pokazuje jak potężną siłą organizującą życie wewnętrzne i myślenie o świecie są pojęcia. Helena nie potrafiła kontrolować swojego zachowania ani zbyt dobrze rozumieć swoich stanów umysłu. Zapamiętała wiele symboli dotykowych, nie wiążąc ich z przedmiotami ani swoimi działaniami. Do przełomowego momentu w jej życiu doszło, gdy stymulowana przez swoją opiekunkę za pomocą dotyku powiązała symbol (serię sygnałów dotykowych) *w-o-d-a* z wrażeniem wywołanym przez polewanie ręki wodą. Skojarzenie pozostałych zapamiętanych symboli z wrażeniami odbyło się szybko. Jak pisze w swojej autobiografii „Każda rzecz miała nazwę, a każda nazwa dawała życie nowej myśli."

A3. Od 1984 roku tworzona jest globalna ontologia projektu CYC (nazwa pochodzi od fragmentu słowa enCYClopedia), w ramach której powstała baza wiedzy zawierającej opisy mi-

lionów pojęć i powiązań pomiędzy nimi. 6000 podstawowych pojęć koniecznych do opisu rzeczywistości jest silnie powiązana z bezpośrednim doświadczeniem bycia w świecie, pozostałe pojęcia są z nich konstruowane (Lenat,1995). Epistemologia konstruktywistyczna ma długą historię i pojawia się w różnych dziedzinach nauk społecznych, w filozofii, badaniach literackich, pedagogice, psychologii, socjologii, a także a architekturze i sztuce. W radykalnej wersji konstruktywizm jako pragmatyczne podejście promował Ernst Von Glasersfeld, definiując dziedzinę we wstępie do książki „Radical Constructivism: A Way of Knowing and Learning" (1995) w następujący sposób: „Radykalny Konstruktywizm to niekonwencjonalne podejście do problemu wiedzy i jej zdobywania. Jego podstawowym założeniem jest stwierdzenie że wiedza, niezależnie od sposobu jej definiowania, jest w głowach ludzi, a osoba myśląca nie ma innego wyboru jak tylko konstruować to co wie na podstawie swojego doświadczenia. […] Doświadczenie i interpretacja języka nie jest tu wyjątkiem".

Wiedza i pojęcia są „w głowie", czyli w mózgu. Procesy zachodzące w mózgach przebiegają drogami wyżłobionymi przez doświadczenie. Zadaniem neurokognitywnej lingwistyki jest zrozumienie tych procesów.

A4. To, co się spontanicznie przejawia w treści naszej świadomości mając decydujący wpływ na nasze zachowanie i wypowiedzi, jest wynikiem trzech zasadniczych czynników:

- **Determinizmu genetycznego**, uwarunkowań tworzących ogólne ramy naszych zdolności przeżywania faktu bycia w świecie, zdolności percepcyjnych, możliwości tworzenia i zapamiętywania skojarzeń, poziomu inteligencji dostępnemu człowiekowi.
- **Determinizmu neuronalnego**, wyniku doświadczeń życiowych, wdrukowania (imprintingu) i wychowania, tworzącego konkretne struktury funkcjonalnych podsieci, możliwości powstawania stanów umysłu w oparciu o substrat, jakim jest mózg.
- **Czynników stochastycznych**, wpływających na zachowanie, które do pewnego stopnia w przypadkowy sposób modyfikują dynamiczne procesy zachodzące w mózgu, wpływając na powstające skojarzenia, tok myślenia i podejmowane decyzje.

Nie możemy myśleć inaczej, niż pozwala nam na to aktywność neuronalna. Używania języka nie można oddzielić od ogólnych procesów myślenia i działania, w mózgu nie mamy jakieś wyodrębnionej podsieci reprezentującej pojęcia, wszystko jest ze sobą silnie sprzężone. Chociaż często konfabulujemy, wymyślając pozornie racjonalne interpretacje swojego zachowa-

nia, prawdziwe przyczyny nie są nam często znane i mogą się okazać zbyt skomplikowane by je odkryć i zrozumieć.

A5. Determinizm na poziomie genetycznym czy neuronalnym nie oznacza wcale, że jesteśmy automatami, niezdolnymi do modyfikacji swoich zachowań, jak jakieś skomplikowane mechanizmy zegarowe. Układy uczące się ciągle się zmieniają pod wpływem działających na nie bodźców. Plastyczność mózgu widoczna jest na każdym etapie, od genów, których ekspresja modyfikowana jest przez czynniki środowiskowe, synaps które zmieniają się w wyniku pobudzenia neuronu, do funkcjonalnych podsieci, które tworzą się gdy uczymy się nowych pojęć lub docierają do nas nowe informacje. Prawa fizyki na poziomie kwantowym są określone w sposób probabilistyczny, a nie ściśle deterministyczny. Poziom, na którym można mówić o zdarzeniach mentalnych jest od poziomu kwantowego niesłychanie odległy. Zdarzenia mentalne powstają w emergentny sposób jako rezultat neurodynamiki materii mózgu, zależnej nie tylko od połączeń i pobudzeń neuronów, całej biochemii organizmu ale i całego środowiska, z którym ten organizm jest sprzężony. W sieci neuronowej sterującej robotem zachodzą procesy fizyczne, ale jeśli robot potrafi się uczyć procesy te nie wyjaśnią w pełni jego zachowania. Żeby zrozumieć reakcje takiego robota trzeba poznać historię jego rozwoju, wiedzieć od kogo i czego się nauczył, a więc opisać procesy nim sterujące nie na poziomie programów sterujących ale wiedzy, która decyduje o jego zachowaniu. Sieć neuronowa robota, podobnie jak mózg człowieka, jest substratem który umożliwia powstanie emergentnych procesów mentalnych. Na razie są one znacznie bardziej złożone u ludzi, ale żadne prawo przyrody nie gwarantuje, że tak będzie w przyszłości. Już teraz na powierzchni 1 cm^2 mieści się 500 mln. tranzystorów, a do końca tej dekady nowe procesy technologiczne pozwolą na budowę struktur przypominających neurony i synapsy o podobnej gęstości jak w mózgach ssaków.

Determinizm neuronalny oznacza tylko tyle, że w danej chwili możliwości mózgu są ograniczone, dziecku nie przychodzą spontaniczne do głowy matematyczne formuły teorii superstrun, ale gdy dorośnie i zostanie ekspertem w tej dziedzinie fizyki będzie to prawdopodobne. Plastycznośc mózgu jest duża ale możliwości zmian są ograniczone, genetyka nie pozwoli nam widzieć kolorów w ultrafiolecie, jak robią to owady, ani odczuwać wibracji pola elektrycznego, jak robią to niektóre ryby. Możliwości skojarzeniowe mózgu i pojemność pamięci roboczej nie pozwolą nam zrozumieć zbyt złożonych zależności.

A6. Nie opiszemy w pełni rzeczywistości korzystając z pojęć, które nie są adekwatne do opisu zachodzących w mózgu procesów. Trzeba zrozumieć, jak kodowane są pojęcia, co ma wpływ na ich rozumienie, jak w systematyczny, chociaż wielce przybliżony sposób można je opisywać tak, by jak najlepiej oddać naturę neurodynamicznych procesów jakie za nimi stoją. Lingwiści specjalizują się w fonetyce, fonologii, morfologii, syntaktyce, leksykografii, ontologiach, semantyce, pragmatyce i innych dziedzinach, ale język zależy od integracji wielomodalnej informacji, łączy się z percepcją i myśleniem. Tylko neuronowe teorie języka mogą prawidłowo opisać wszystkie jego aspekty, metody formalne nie opisują dobrze dynamicznych funkcji języka (Spivey, 2007). Przetwarzanie informacji w sieciach neuronowych (elementy takich sieci reprezentują modele neuronów) lub sieciach koneksjonistycznych (elementy takich sieci reprezentują pojęcia a ich połączenia relacje pomiędzy nimi) daje psychologicznie interesujące rezultaty, a więc złożoność mózgu nie jest głównym problemem (Duch i inn. 2008)! Celem informatyki neurokognitywnej jest tworzenie i badanie uproszczonych modeli wyższych czynności poznawczych, myślenia, rozwiązywania problemów, uwagi, kontroli zachowania, świadomości, języka, pomagając w lepszym zrozumieniu procesów za nie odpowiedzialnych (Duch, 2009). Wyników tych badań są inspiracją do tworzenia praktycznych algorytmów komputerowych.

Następnym rozdział zawiera ogólną dyskusję na temat reprezentacji pojęć w mózgu, rozdział trzeci przedstawia nieco wyników eksperymentalnych i modeli komputerowych, a rozdział czwarty omawia bardziej szczegółowo rezultaty naszych własnych eksperymentów. konsekwencje i zawiera nieco spekulacji. Piąty rozdział zawiera uwagi na temat kompromisu pomiędzy plastycznością i stabilnością na różnych poziomach.

2. Pojęcia i mózgi – rozważania ogólne

B1. Dobrym wprowadzeniem do zagadnień związanych z naturalizacją epistemologii, reprezentacją pojęć w mózgu i tworzeniem się przekonań i teorii są książki Paula Churchlanda (1989, 1995), przedstawiające podstawowe idee dotyczące modeli neuronowych. Symboliczny, werbalny opis stanów umysłu napotyka na liczne trudności (Schwitzgabel, 2011), związane między innymi z tym, że rozkład aktywności grup neuronów w mózgu zmienia się w sposób ciągły. Mamy więc nieskończenie wiele stanów mózgu i tylko najczęściej powtarzającym

się kategoriom tych stanów możemy przypisać symboliczne nazwy. Anderson (2010) dokonał meta-analizy 165 eksperymentów z użyciem neuroobrazowania dotyczących różnych funkcji związanych z używaniem i rozumieniem mowy i tekstów czytanych. Prawie wszystkie pola Brodmanna (reprezentujące zlokalizowane obszary kory mózgu) były ze sobą funkcjonalnie powiązane, tzn. wykazywały skoordynowaną aktywność w analizowanych eksperymentach. Żadna inna funkcja nie angażowała tak rozległych funkcjonalnych sieci mózgu jak zadania językowe.

B2. Nie potrafimy przypisać nazw nietypowym pobudzeniom mózgu zachodzącym w czasie snu, halucynacji, czy narkozy. Brak pobudzenia kory wzrokowej V1, przy jednoczesnej aktywacji obszaru MT i zakrętu skroniowego dolnego (IT) może być interpretowane jako wrażenie ruchu rozpoznanej postaci bez wrażeń wzrokowych (Duch, 2011). Tylko niektóre konfiguracje jednocześnie pobudzonych obszarów mózgu mają na tyle jednoznaczną interpretację by przypisać im symboliczną nazwę, czyli reprezentację fonologiczną, związaną z aktywacją kory słuchowej w obszarze górnego zakrętu skroniowego (Okada, Hickok, 2006). Rozkład aktywacji pól mózgu $a(x,t)$ można opisać za pomocą modelu sieciowego, reprezentując każde pole przez grupę współpracujących, a jednocześnie konkurujących ze sobą neuronów. Model wektorowy jest nieco prostszy, zapisuje zbiór współczynników aktywacji $a(x_i,t)$ uśrednionych dla intersujących regionów (ROI, Regions of Interest), czyli pól neuronowych położonych koło x_i. W modelu sieciowym można śledzić zmiany aktywności elementów, procesy synchronizacji i desynchronizacji neuronów. Model wektorowy ma większe trudności z uchwyceniem dynamicznej natury powstawania aktywacji w mózgu, prezentując uśrednione prototypy dla wybranych aktywacji, podobnie jak to widać w neuroobrazowania, przedstawiającym uśrednioną aktywację różnych obszarów mózgu. Jednak również w modelu wektorowym można się pokusić o opisanie całej trajektorii zmian stanu sieci.

B3. Model sieciowy czy też wektorowy jest bliższy rzeczywistości fizycznej, chociaż jego interpretacja, wiążąca symbole ze stanami mózgu, może początkowo nastręczać trudności. Można w nim rozróżnić dwie podstawowe składowe: reprezentację formy (związanej z fonologią i ortografią), oraz reprezentację semantyki, związanej z rozkładem pobudzeń pozostałych obszarów mózgu. W rzeczywistości mamy do czynienia z jednym niepodzielnym stanem mózgu, ale skojarzenia na poziomie fonologicznym i na poziomie semantycznym można do pewnego stopnia rozpatrywać niezależne. Fonologiczne reprezentacje są funkcją zakrętu

skroniowego górnego (Okada, Hickok, 2006). Czysto semantyczne skojarzenia mogą być różnorodne, w zależności od kontekstu aktywność różnych obszarów mózgu może otworzyć drogę prowadzącą do kolejnego stanu.

B4. Rozumienie pojęć na poziomie koncepcyjnym sprowadza się do próby zdefiniowania danego pojęcia za pomocą szeregu innych pojęć w nadziei, że pojęcia definiujące będą bardziej zrozumiałe. Steven Harnad (1990) zainicjował interesująca dyskusję na temat nabierania znaczeń przez symbole (*the symbol grounding problem*) w systemach formalnych. Po 15 latach podsumowanie tej dyskusji (Taddeo i Floridi, 2005) doprowadziło do wniosków, że do pojawienia się sensu autonomiczny agent (program lub robot) musi przypisać symbole do sensomotorycznych interakcji ze środowiskiem, mieć zdolności do tworzenia reprezentacji stanów środowiska, rozróżniania kategorii tych stanów, oraz komunikacji z innymi agentami by skoordynować znaczenie używanych symboli. Symbole są intencjonalne, odnoszą się do świata zewnętrznego, do możliwości percepcji i działania w świecie, wiedzy o tym, czego można się spodziewać po zewnętrznym środowisku. Stany wewnętrzne robota czy zwierzęcia nie dadzą się w pełni opisać za pomocą skończonej liczby symboli. Każde pojęcie odnosi się do klasy dość zróżnicowanych stanów wewnętrznych, które mogą powstać na skutek rozpoznania obiektu, obserwacji zdarzenia lub skojarzeń myślowych.

B5. Świadomość, jak pisał już John Locke, jest postrzeganiem stanów swojego umysłu. Pojęcia odpowiadające fizycznym obiektom i zdarzeniom wydają się całkiem odmienne niż pojęcia abstrakcyjne, ale w każdym przypadku jest to kategoryzacja stanu mózgu skojarzona z symboliczną etykietą, wewnętrzne postrzeganie i komentowanie powstających stanów. Patrząc na stojącą przede mną butelkę z zielonego szkła z czystą wodą mam liczne wrażenia wzrokowe związane z kształtami i kolorami. Przekonanie, że świat tak właśnie wygląda jest jednak złudne, bo doświadczam w świadomy sposób jedynie perceptów powstałych z przetworzonych w bardzo skomplikowany sposób bodźców sensorycznych. Obraz padający na siatkówkę może się w dużych granicach zmieniać, ale wrażenie koloru pozostaje stałe. Podstawowe dźwięki mowy rozpoznawane są jednoznacznie pomimo różnic w szybkości mówienia czy wysokości głosu. Mózg w toku ewolucji nauczył się przetwarzać wiele często powtarzających się bodźców tak, by wrażenia były stałe. Ułatwia to podejmowanie świadomych decyzji w sytuacjach zbyt zróżnicowanych by je w pełni zinternalizować i podejmować decyzję w nieświadomy sposób. Czy owoc jest już dojrzały i warto podjąć wysiłek by go zerwać?

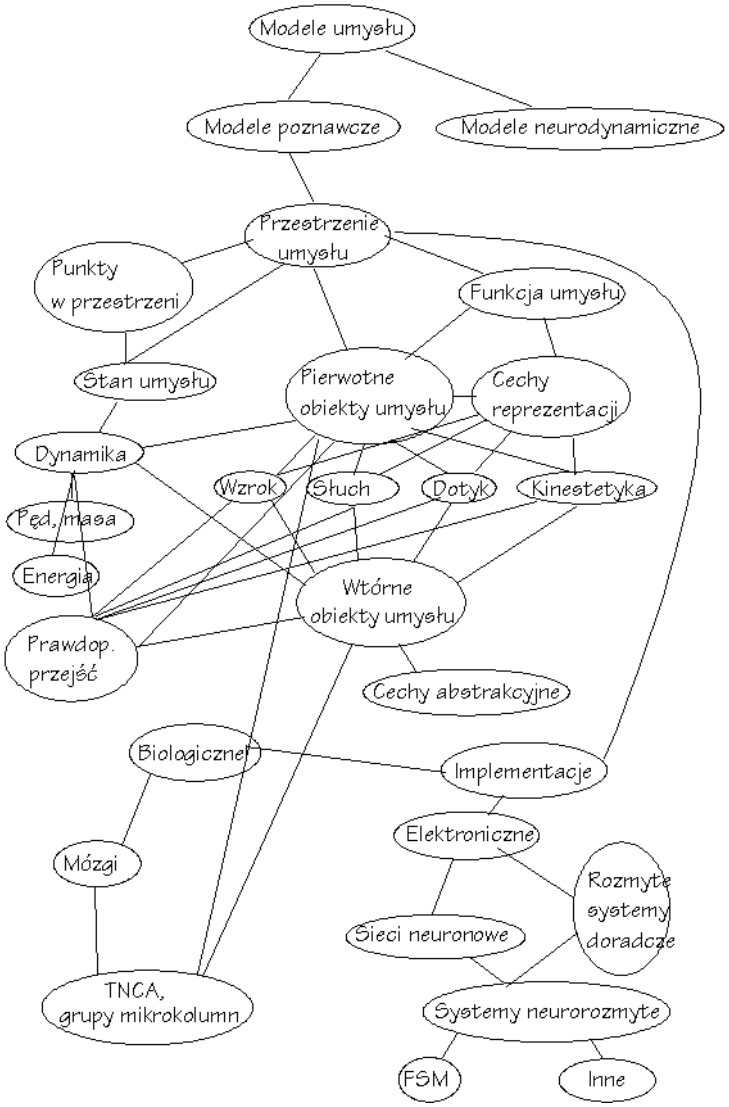
Czy słyszane dźwięki składają się w zrozumiałe słowo? Konteksty nawet tak prostych zadań są zbyt złożone by można je było zautomatyzować w postaci nieświadomych decyzji.

B6. Aktywacja różnych obszarów mojego mózgu w przypadku stojącej obok butelki soku wiśniowego jest inna, kształt i kolor całkiem odmienne od zielonej butelki, jednak obydwu wrażeniom przypisuję to samo pojęcie *butelka*. Pojęcie nie wskazuje na samo wrażenie, czyli jednoznacznie określony stan kory zmysłowej w mózgu, ale na całą klasę dość odmiennych stanów. Informacja, która pozwala mi klasyfikować wrażenia opisując je za pomocą pojęć, jest wynikiem automatycznego sposobu przetwarzania, tworząc w kolejnych etapach niezmiennicze reprezentacje i abstrakcyjne uogólnienia istotnych cech obiektów. Kategorie naturalne nie są ostro określone i powstają przez generalizację prototypów złożonych wrażeń. Reprezentacje wewnętrzne nie musza odnosić się do cech obiektów, wystarczy informacja o podobieństwie powstałego stan aktywacji mózgu do poprzednio zapamiętanego stanu. Symbol związany z danym obiektem potrafi nie tylko przywołać jego wyobrażenie, odtwarzając z grubsza stan kory zmysłowej, ale przygotowuje mózg do określonego działania.

Ugruntowanie sensu pojęć w działaniu w świecie, zgodnie z projektem enaktywizmu podkreślającego rolę środowiska w procesie samoorganizacji organizmu i powstania umysłu (Varela i inn. 1991; Barsalou 2008) nie może się ograniczać do działania fizycznego. Różnica pomiędzy procesami w mózgu związanymi z działaniami fizycznymi i mentalnym polega głównie na słabej aktywacji pierwotnej kory ruchowej i mięśni dla procesów mentalnych. W obu przypadkach aktywacje obejmują większą część mózgu, który na poziomie świadomym, postrzeżenia wewnętrznego, ma tylko pośrednią wiedzę o zewnętrznym otoczeniu, odczytaną ze stanu kory analizujące dane zmysłowe.

Rozumienie pojęć abstrakcyjnych jest związane bardziej z przygotowaniem skojarzeń, które pozwalają je użyć w procesach myślowych, niż z działaniami fizycznymi. Tylko niewielka część pojęć nabiera sens przez ugruntowanie ich znaczenia w relacjach sensomotorycznych. Większość to pojęcia oderwane od bezpośredniego doświadczania i działania w świecie. W raporcie z 1994 roku nazwałem takie pojęcia „wtórnymi obiektami umysłu" (Duch, 1994), w odróżnieniu od obiektów pierwotnych, których wewnętrzne reprezentacje opierają się na aktywacji kory przetwarzającej informacje z różnych zmysłów oraz kory ruchowej. Takie stany próbowałem opisywać w „przestrzeniach umysłu" (mind spaces), starając się zdefiniować przestrzeń stanów psychologicznych i opisać stan wewnętrzny jako punkt w takiej przestrze-

ni. W odróżnieniu od opisu stanu aktywacji różnych obszarów mózgu, do czego trzeba stosować modele opisujące neurodynamikę, miałem nadzieję, że procesy mentalne da się rozpatrywać w przestrzeniach umysłu, których poszczególne wymiary będą miały znaczenie zrozumiałe dzięki introspekcji, będą opisywane przez mierzalne cechy jak i specyficzne jakości wrażeń (qualia) odnoszące się do stanu kory zmysłowej. Geometryczny model procesów mentalnych powinien być bardziej zrozumiały, bliższy perspektywie wewnętrznej, niż modele neuronowe (Duch 1997, 2002, 2002a). Umysł można w tym ujęciu uznać za cień neurodynamiki, która jest znacznie bogatsza. W lingwistyce teorię przestrzeni mentalnych rozwijał Fauconnier (1994,1997), oraz kognitywistyczny model przestrzeni konceptualnych opisał Gärdenfors (2000), jednakże modele te nie próbowano powiązać z neurodynamiką.



Rys. 1 Siatka pojęć reprezentacji mentalnych i ich modeli.

Teoretycznie można dokonać matematycznej transformacji $U_i(t)=T(A_j(t))$ aktywacji różnych obszarów mózgu $A_j(t)$, obliczając na tej podstawie cechy wrażeń zmysłowych oraz emocji składające się na stan umysłu $U_i(t)$. Zdarzenia mentalne można by było wówczas opisać przez trajektorie w przestrzeni psychologicznej. Chociaż zastosowania tego pomysłu do opisu kategoryzacji obiektów na podstawie kilku cech, takich jak kolor, wielkość i kształt, wyglądały obiecująco i udało się znacznie uprościć opis neurodynamiki za pomocą modeli neurorozmytych (Duch, 1996, 1996a) jest tu kilka trudności. Przestrzeń mentalna ma wiele wymiarów, więc trudno sobie wyobrazić zachodzące w niej procesy. Z matematycznego punktu widzenia mamy do czynienia z przestrzeniami o zmiennej liczbie wymiarów, w których da się zdefiniować relacje podobieństwa. Nie potrafimy dobrze opisać swojego stanu wewnętrznego (Schwitzgabel, 2011), więc nie bardzo wiemy jakie wymiary powinny się w takich przestrzeniach znaleźć. Skojarzenia w mózgu powstają przez podobieństwo rozkładów aktywności, lub podobieństwa wyższego rzędu, które nie dają się sprowadzić do prostych własności. Wymiary w przestrzeni umysłu musiałyby więc reprezentować nie tylko wrażenia i cechy, ale i podobieństwa. Teoria integracji lub mieszania pojęć (*conceptual blending*), rozwinięta przez Gillesa Fauconniera i Marka Turnera (2002), uznawana za ogólną teorię zjawisk poznawczych, zakłada aktywację fragmentów wcześniej zapamiętanych doświadczeń i relacji, z których tworzą się nowe wyobrażenia i opisujące je pojęcia. Geometryczny opis tego procesu musi wykorzystywać wymiary oparte na podobieństwie.

B7. Paul Churchland (1989,1995) naszkicował w swoich książkach sposób tworzenia się zbioru przekonań i rozumienia relacji pojęciowych w wyniku uczenia się. W nieco zmodyfikowanej wersji można go przedstawić następująco: przekonania są wynikiem procesów skojarzeniowych pomiędzy konfiguracjami możliwych (potencjalnych) pobudzeń sieci neuronów. Skojarzenia te obejmują również układ ruchowy, stwarzając predyspozycje do pewnych działań, wypowiedzi lub zmiany stanu mózgu na kolejny stan z nim skojarzony. Z wieloma lokalnymi rozkładami aktywności można związać aktywacje fonologicznych reprezentacji pojęć językowych, czyli symboli wskazujących na właściwy sens semantyczny danej aktywacji. Teoria tworzy się w mózgu stopniowo w czasie nauki, zmiany możliwe są dzięki plastyczności (zdolności do zmian) sieci neuronowych. Emocje mogą tu bardzo pomagać, bo zwiększa-

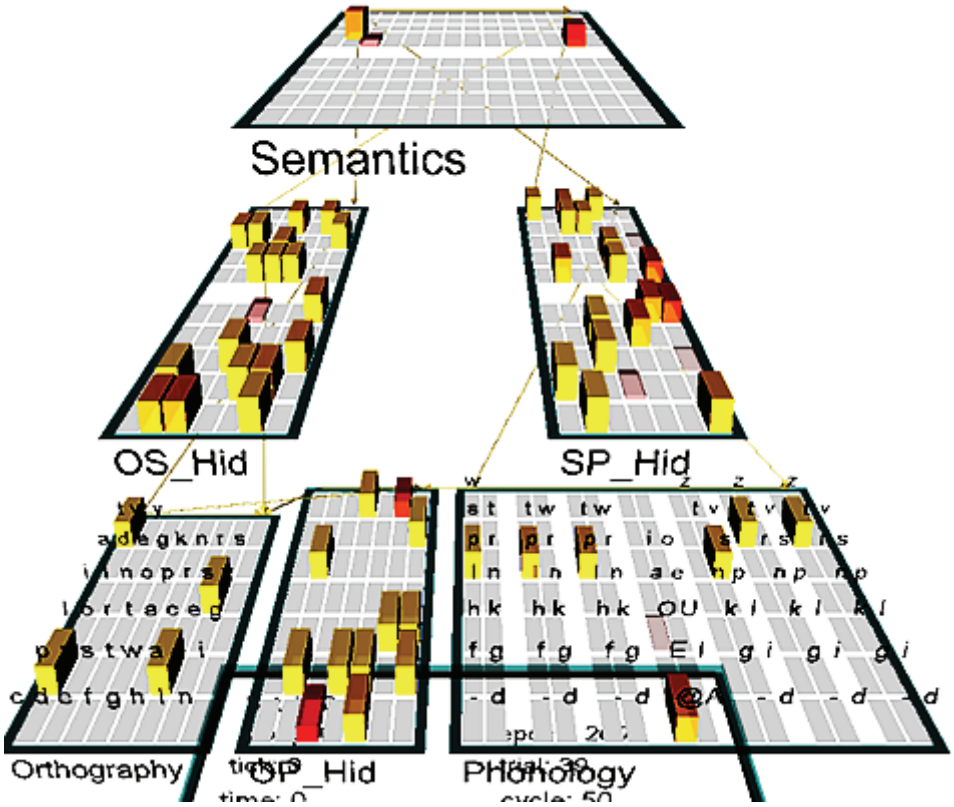
jąc gęstość neuromodulatorów ułatwiają fizyczne zmiany zachodzące w strukturze sieci w czasie uczenia. Istnieją tu znaczne różnice indywidualne, skorelowane ze zdolnościami do uczenia się. Jeśli zmiany w danym mózgu są możliwe tylko na skutek gwałtownego pobudzenia emocjonalnego, następuje wdrukowanie przeżywanych w tym czasie informacji, uformowanie się nowych ścieżek pobudzeń, a po opadnięciu emocji niezdolność do racjonalnej oceny faktów, wiara w teorie spiskowe. Model wewnętrzny jest rzadko w pełni werbalizowalny, powstaje zbiór wzajemnie ze sobą skojarzonych stanów, więc na poziomie symbolicznym przekonania nie muszą sprowadzać się do zbioru spójnych reguł.

B8. Rodney Brooks wprowadził artykułem „Słonie nie grają w szachy" (Brooks, 1986) nowe tendencje rozwojowe w robotyce, odrzucając pomysły budowania symbolicznych modeli umysłu opartych na wewnętrznych reprezentacjach stanów środowiska. Inteligentne zachowania prostych organizmów nie wymagają od nich przechowywania wewnętrznych reprezentacji, chyba że za takie uznać zbiór zależnych od kontekstu reakcji. Brooks próbował pokazać, że można stworzyć inteligentnego robota formując stopniowo funkcje jego mózgu, modelowane za pomocą sieci neuronowych, w naturalny sposób przez interakcję z otoczeniem, podobnie jak rozwija się mózg dziecka. Był to zasadniczy odwrót od wcześniejszych prób budowy sztucznej inteligencji na czysto logicznych podstawach, przy całkowitym ignorowaniu biologii. Głównym projektem, mającym udowodnić słuszność takiego podejścia, miała być budowa robota o nazwie Cog (Brooks i Stein, 1994). Próba ta nie zakończyła się jednak powstaniem umysłu na wzór ludzki, a jedynie zbiorem odruchów pozwalających na proste reakcje. Był to ciekawy eksperyment pomagający wyznaczyć granice pomiędzy wrodzonymi a wyuczonymi umiejętnościami. Jakiejś formy reprezentacji nie da się jednak uniknąć. Pomimo tego robotyka rozwojowa stała się obecnie bardzo ważną dziedziną, powiązaną w psychologią rozwojową, a nadzieje na rozwój bardziej złożonych form poznania i działania na tej drodze nadal się utrzymują.

3. Reprezentacja pojęć

C1. Stany mózgu i ich aproksymacje opisane zostały w pracy (Duch, 2010). Nie mamy na razie dostatecznie dobrych metod obserwacji by śledzić zmiany aktywacji wielu obszarów mózgu z duża rozdzielczością czasową i przestrzenną. Połączenia pomiędzy poszczególnymi

obszarami mózgu znane są z dość słabą rozdzielczością rzędu 100 obszarów, ale w ramach projektu ludzkiego konektomu planowane jest zwiększenie liczby interesujących regionów (ROI, Regions of Interest) o rząd wielkości. Pojęcia reprezentowane przez kwazi-stabilny stan rozkładu aktywności tych obszarów można częściowo opisać przez podobieństwo rozkładów obrazujące sąsiedztwo i relacje z innymi pojęciami, synonimami, antonimami. Techniki neuroobrazowania (fMRI, PET) dają obrazy uśrednione w czasie kilku sekund lub dłuższym, ale ich rozdzielczość jest bardzo duża, rzędu kilkudziesięciu tysięcy obszarów o boku 1-3 milimetrów. Rozdzielczość czasowa elektroencefalografii (EEG) lub magnetoelektroencefalografii (MEG) jest rzędu milisekund, ale lokalizacja źródeł sygnału jest tylko rzędu centymetrów. Podobne ograniczenia ma spektroskopia w bliskiej podczerwieni (NIRS-OT). Niemniej nawet przy tak ograniczonych metodach obserwacji można już wiele powiedzieć o reprezentacji pojęć w mózgu. W innych dziedzinach neuronauk poznawczych możliwe są badania na zwierzętach, ale niewiele nam mogą one powiedzieć o języku.



Rys. 2 Sieć neuronowa, której elementy (neurony) reprezentują ortografie, fonologię i semantykę.

C2. Synchroniczna aktywności grup neuronów (NCA, *neural cell assemblies*) jest dobrą podstawą do modelowania procesów neurolingwistycznych (Damasio i inn. 1996;Pulvermuller, 2003; Dehaene i inn. 2005), uwzględniając ogólne mechanizmy działania pamięci (Lin, Osan, Tsien, 2006). Modele uwzględniają zwykle warstwę fonologiczną, ortograficzną i semantyczną (Rys. 2). Pierwsze dwie warstwy związane są z reprezentacją symboliczną pojęcia, przypisaniem mu nazwy, a warstwa semantyczna ma reprezentować wszystkie pozostałe obszary mózgu, w których pojawia się aktywacja w wyniku prezentacji danego pojęcia. Parametry modelu zmieniają się w procesie uczenia tak, by po pokazaniu słowa w formie pisanej, np. *flag*, w warstwie fonologicznej pojawiły się aktywacje pozwalające na jego poprawną wymowę, a w warstwie semantycznej pobudziły się jednostki odpowiadające cechom, jakie danemu pojęciu można przypisać. Pomiędzy każdą parą warstw są grupy neuronów pośredniczących, przetwarzających sygnały w taki sposób, by po pokazaniu danego pojęcia w jednej z form – pisanej, fonologicznej lub semantycznej –pobudzając jedna warstwę uzyskać właściwe pobudzenia pozostałych dwóch warstw. Model przedstawiony na rysunku zrealizowany został w symulatorze Emergent (Aisa i inn. 2008; O'Reilly i Munakata, 2000), który pozwala na użycie uproszczonych, ale dobrze umotywowanych biologicznie, modeli neuronów i algorytmów odpowiedzialnych za uczenie korelacyjne (Hebbowskie) i uczenie wykonywania zadań. Warstwa semantyczna liczy sobie tu 140 elementów (neuronów), a wysokość słupków na rysunku 2 reprezentuje aktywność grup neuronów kodujących odpowiednie cechy.

C3. Możliwe jest też badanie znacznie bardziej szczegółowych modeli, np. w modelu Garagnani i inn. (2009) uwzględniono interakcje 6 obszarów: pierwotnej kory słuchowej, pasa słuchowego, obszaru Wernickiego, boczno-brzusznej kory przedczołowej, kory przedruchowej Broki, oraz pierwotnej kory ruchowej. Reprezentacja pojęć tworzy się w tym modelu spontanicznie w postaci silnie połączonych mikroobwodów znajdujących się w anatomicznie różnych obszarach realizujących funkcje postrzegania-działania (Pulvermuller, 2003). W takim modelu możliwa jest aktywacja kilku reprezentacji leksykalnych jednocześnie, podczas gdy model warstwowy, przedstawiony na rysunku 2, ma tendencję do szybkiego przeskakiwania od jednej reprezentacji do drugiej. Więcej szczegółów na temat neuroanatomii obszarów zaangażowanych w funkcje językowe znaleźć można w pracy (Duch, 2010).

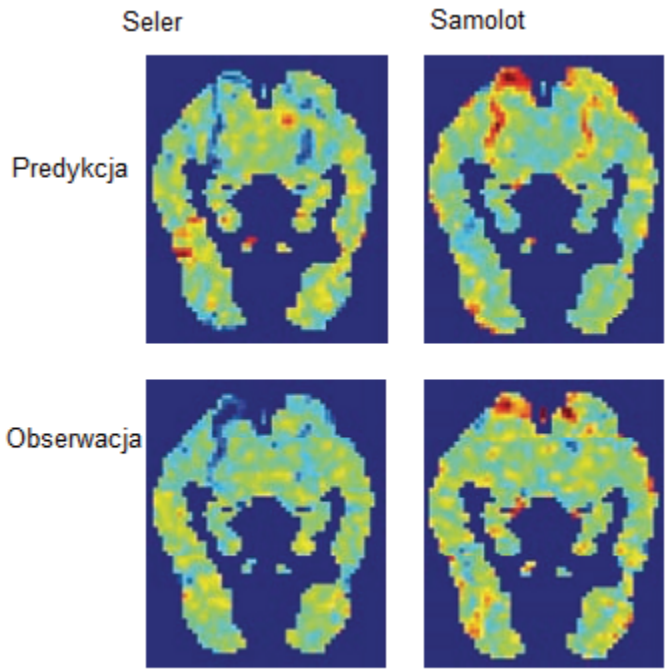
C4. Czy możemy zobaczyć reprezentacje pojęć w mózgu? Podglądanie intencji i rozpoznawanie stanów mentalnych za pomocą metod obrazowania jest już dość zaawansowane (Hay-

nes i Rees, 2006; Haynes i inn. 2007). Coraz więcej eksperymentów używających metod neuroobrazowania pokazuje obrazy rozkładu aktywacji obszarów mózgu w czasie myślenia o jakimś pojęciu, inicjowanego przez obraz, usłyszaną nazwę, napisaną nazwę, wybór jednego z kilku pojęć w myślach. Pomimo indywidualnych różnic aktywacje u różnych ludzi są na tyle podobne, że prosty klasyfikator liniowy może się nauczyć je rozróżniać i przewidywać jak będą wyglądać dla nowych pojęć (Mitchell i inn. 2008). Żeby to zrobić musimy najpierw zamienić pojęcie *P* na wektor je opisujący V(*P*), a następnie nauczyć się korelacji pomiędzy tym wektorem i macierzą M(*P*) przechowującą aktywność tysięcy wokseli mózgu. Następnie na znanych przykładach próbujemy określić parametry *W* transformacji M(*P*)=F(V(*P*);*W*) pomiędzy wektorem V(*P*) i macierzą M(*P*) dla kilkudziesięciu lub więcej pojęć *P*. Mając taką transformację możemy przewidzieć jaki będzie rozkład pobudzeń dla nowych pojęć, dla których nie wykonano jeszcze eksperymentu.

C5. Jak utworzyć wektory semantyczne V(*P*)? W pracy (Mitchell i inn. 2008) wybrano 25 cech semantycznych, które odnoszą się do postrzegania i działania: czy pojęcie *P* związane jest z widzeniem, słuchaniem, wąchaniem, smakowaniem, dotykaniem, strachem, jedzeniem, mówieniem, poruszaniem, popychaniem, pocieraniem, bieganiem, podnoszeniem, jechaniem, noszeniem na sobie, czyszczeniem, wypełnianiem, otwieraniem, połamaniem … Dla obliczania wektorów semantycznych użyto bardzo dużego korpus tekstów (rzędu biliona słów), dla którego policzono korelacje danego pojęcia *P* z występowaniem tego typu cech w jego pobliżu. W ten sposób składowe wektora V(*P*) określają korelacje każdej z 25 cech z danym słowem. Samolot można widzieć i słyszeć ale nie smakujemy go ani nie wąchamy, nie nosimy na sobie, tylko nim jeździmy. Tak przygotowane wektory V(*P*) i obrazy fMRI mózgów zamienione na macierze M(*P*) pozwalają na nauczenie się parametrów *W* transformacji M(*P*)= F(V(*P*);*W*) na skanach mózgu wykonanych dla kilkudziesięciu pojęć. Znając parametry transformacji można dla nowego pojęcia *P'* utworzyć najpierw wektor V(*P'*) obliczając jego korelację z wybranymi cechami w korpusie tekstów, a potem obliczyć M(*P'*)= F(V(*P'*);*W*). Wykonując kolejne eksperymenty można sprawdzić na ile przewidywania zgadzają się z obserwacjami.

Pomimo dużych uproszczeń takiego podejścia wyniki są intersujące: wybrano 12 kategorii semantycznych (zwierzęta, części ciała, budynki, ich elementy, ubrania, meble, przybory kuchenne, narzędzia, insekty, warzywa, pojazdy, obiekty stworzone przez ludzi), a w każdej z

tych kategorii 5 pojęć. Obraz fMRI utworzono z uśrednienia 6 prezentacji podpisanych obrazków dla każdej z 9 osób biorących udział w eksperymencie. Dla 60 znanych skanów klasyfikator uczono na 58 z ich i przewidywano 2 pozostałe, kolejno wymieniając skany do uczenia i testu tak, by ocenić dokładność przewidywania dla wszystkich 60 z nich. Przewidywany rozkład M(P) porównywano z wszystkimi zapisywanymi. W 72-85% procentach przypadków dla różnych uczestników badania przewidywany rozkład przypominał najbardziej rzeczywisty (przykład na rys. 3). Te różnice wyników dla poszczególnych osób wiążą się z rozmywaniem obrazu na skutek ruchów głowy w czasie eksperymentów. Woksele, które są dobrze przewidywane leżą głównie w lewej półkuli w obszarach powiązanych z rozpoznawaniem obiektów wzrokowych (zakręt dolnoskroniowy, kora ciemieniowa, zakręt wrzecionowaty), korą ruchową, bruzdą śródciemieniową, korą oczodołową. Ekstrapolacja do nowych kategorii, dla których nie było żadnych przykładów obniżyła trochę dokładność do 64-78% dla indywidualnych uczestników.

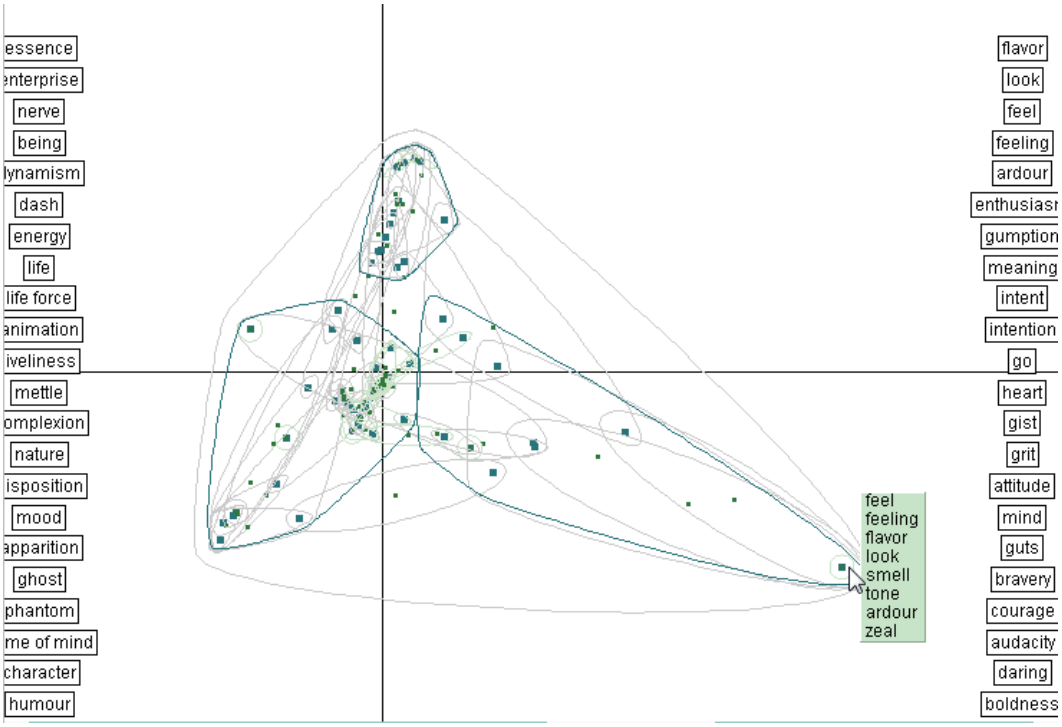


Rys. 3 Przewidywane i obserwowane rozkłady aktywności neuronów dla dwóch rzeczowników.

C6. Pobudzenia mózgu nadają się więc jako naturalna baza reprezentacji semantycznych. Nie należy się spodziewać doskonałej korelacji, bo różne grupy neuronów, zależnie do kontekstu, mogą być zaangażowane w różnym stopniu w wykonywanie tej samej funkcji. Kontekst, w

którym pojawia się jakieś pojęcie, zupełnie zmienia te rozkłady. Dla przykładu, słówko „spirit" wymienione jest w 79 kontekstach, które można pogrupować w 69 klik (http://dico.isc.cnrs.fr/en/index.html), reprezentujących synonimy określające pojęcia, różne znaczenia danego słowa. Hierarchiczne uporządkowane kliki grupujące synonimy dla różnych znaczeń tego pojęcia pokazane są na rysunku 4. Ten obraz pasuje dobrze do hierarchicznej i modularnej organizacji sieci neuronowych w mózgu (Meunier i inn. 2010).

Podobieństwo fonologiczne i semantyczne pomiędzy słowami może prowadzić do podobnych aktywacji mózgu dla wybranego, izolowanego pojęcia, ale w większości przypadków torowanie związane z historią i kontekstem całkowicie ujednoznacznia interpretacje sensu słów. Jest za to odpowiedzialna konkurencja pomiędzy procesami neuronowymi. Mechanizm znany pod nazwą „zwycięzca bierze wszystko" (WTA, winner-takes-all) polega na tym, że po osiągnięciu pewnego progu synchronizacji neuronów ich aktywność zahamuje konkurencję (O'Reilly, Munakata, 2000). W efekcie „do głowy przychodzi" tylko właściwy sens danego pojęcia w danym kontekście.



Rys. 4 Reprezentacja różnych znaczeń słowa „spirit" w atlasie semantycznym.

C7. Synchronizacja neuronów zachodzi szybko i stany mózgu zmieniają się prawie skokowo, od jednego rozkładu do drugiego. Nie dotyczy to tylko procesów myślenia, automatyczna segmentacja to podstawa percepcji, ułatwiająca planowanie, zapamiętywanie i łączenie informacji ze sobą. Przejścia pomiędzy poszczególnymi segmentami wewnętrznych doświadczeń zachodzą w wyniku obserwacji istotnych zmian sytuacji, pojawienia się nowych postaci, nowych wątków w interakcji ze światem, nowego miejsca i możliwych celów działania. Do pewnego stopnia przypomina to sekwencje filmu, chociaż przejścia są częstsze i niezauważalne. Świat naszych przeżyć jest sekwencją scen, a stany przejściowe nie są postrzegane (Zacks i inn, 2010). Pomimo różnic szczegółów wynikających z kontekstu można w obrazowaniu za pomocą fMRI dostrzec w czasie słuchania czytanej historii aktywacje mózgu, który reaguje na zmiany sytuacji, umiejscowienie postaci względem siebie i elementów sceny, cele, przyczyny, zmiany czasu i miejsca (Speer i inn, 2009) . Te reakcje podobne są do tych, jakie pojawiają się przy samodzielnym wykonywaniu podobnych czynności, lub przy obserwowaniu jak ktoś inny je wykonuje. Rozumienie sensu jest pewnego rodzaju symulacją sytuacji dokonywaną w wyobraźni. Ważną rolę pełni tu system neuronów lustrzanych.

C8. Czy myślenie jest w pełni zależne od języka czy też może operować od niego niezależnie? Być może język jest konieczny tylko dla przedstawienia problemu, wprowadzenia go do mózgu, a właściwe operacje wnioskowania odbywają się w sposób nieświadomy. Monti i inn. (2009) pokazali, że wnioskowanie logiczne i wnioskowanie oparte na argumentach językowych to różne funkcje mózgu, angażujące odmienne struktury. Argumenty logiczne typu: jeśli zarówno X i Z to nie Y, lub jeśli Y to ani nie X ani nie Z, angażują płaty czołowe i przedczołowe. Argumenty lingwistyczne typu: rzecz X, którą Y widział jak Z brał, lub Z był widziany przez Y biorąc X angażują obszary językowe wokół bruzdy Sylwiusza oraz jądro ogoniaste. W obu przypadkach pojawiają się też aktywacje w licznych polach płatów czołowych i ciemieniowych. Z tych badań wynika, że abstrakcyjne myślenie logiczne i myślenie oparte na pojęciach języka to w znacznym stopniu odmienne procesy.

W przeglądowej pracy Monti i Osherson (2011, w druku) uzasadniają, że rola języka w rozumowaniu dedukcyjnym jest ograniczona do początkowego etapu w którym werbalnie prezentowana informacja ulega zakodowaniu w postaci niewerbalnych reprezentacji. Te reprezentacje są wykorzystywane przez operacje mentalne w oderwaniu od neuronalnych mechanizmów związanych z językiem. Badania procesów wnioskowania prowadzą często do kon-

trowersyjnych wniosków, dlatego Monti i Osherson bardzo dokładnie rozważyli różne formy wnioskowania, starając się pokazać tego przyczyny. Ich konkluzje związane są prawdopodobnie ze stosunkowo prostą naturą problemów rozważanych w eksperymentach. Trudno jest oddzielić aktywność warstwy semantycznej od fonologicznej i ortograficznej. Forma reprezentacji może być drugorzędna, ale pomaga przywołać właściwy stan mózgu. Używanie symboli i ich manipulacja na kartce papieru pozwala zrobić krok rozumowania odrywając się daleko od początkowego opisu. Pojęcia i symbole mogą nie być bezpośrednio zaangażowane w sam proces skojarzeniowy, ale ich rola w ułatwieniu funkcjonowania pamięci roboczej, jak i pamięci zewnętrznej jest kluczowa.

C9. Dijksterhuis i Nordgren (2006) pracujący w Nijmegen Unconscious Laboratory w Holandii doszli do zaskakujących wniosków: decyzje w prostych sprawach podejmowane świadomie są lepsze, a decyzje w sprawach skomplikowanych, w przypadku wielu sprzecznych ze sobą kryteriów, jak to ma miejsce przy wyborze partnera życiowego czy zakupie domu, lepiej podejmować opierając się na intuicji, a nie na racjonalnym wnioskowaniu. Jak wynika z ich doświadczeń ludzie, którzy próbują podejmować racjonalne decyzje w sytuacji, w której nie ma jednoznacznego najlepszego wyboru są ze swoich decyzji mniej zadowoleni niż ludzie podejmujące je intuicyjnie. Większość wnioskowań, zwłaszcza kreatywnych pomysłów, wymaga nieświadomego myślenia. Reakcje intuicyjne opierają się na całościowych ocenach, podobieństwie i emocjach. Skoro nie da się optymalnie podjąć decyzji lepiej zdać się na emocje, bo to one decydują o zadowoleniu z dokonanego wyboru w przyszłości.

Myślenie pojęciowe radykalnie zmienia sposób działania mózgu, umożliwiając powstanie pamięci semantycznej, ale nie gwarantuje optymalnego wyboru jeśli jest wiele sprzecznych ze sobą kryteriów, bo wówczas każda decyzja jest kompromisem. W skomplikowanych sytuacjach argumenty logiczne okazują się często błędne podczas gdy proste rozumowanie heurystyczne oparte na intuicji działa zaskakująco dobrze (Gigerenzer 2009). Wybór intuicyjny opiera się na doświadczeniu, podobieństwie do wcześniej spotkanych sytuacji, przyciągającym trajektorię aktywności neuronów do jakiegoś atraktora.

C10. Pojęcia służą nam do komunikacji, symbole powinny w nas wzbudzać odpowiednie reprezentacje semantyczne. Nie można jednak wzbudzić czegoś, czego nie ma. Jeśli ktoś nie zna smaku duriana to nie da się mu wytłumaczyć, jak on smakuje. Pojęcia tego rodzaju mogą jedynie wskazać na znane doświadczenie. Dzielimy z innymi przedstawicielami naszego ga-

tunku bardzo wiele wspólnych cech, chociaż kulturowe idiosynkrazje powodują czasami nieporozumienia. Czy można werbalnie opisać znane obiekty tak, by je dobrze rozróżnić?

Postawiliśmy sobie (wraz ze studentami, M. Gawarkiewiczem, P. Olszakiem i B. Sikorskim) stosunkowo proste zadanie. Korzystając z dostępnych w Internecie informacji o rasach psów (baz danych, stron związków kynologicznych, opisów encyklopedycznych) zbieraliśmy szczegółowe informacje o ich własnościach. Czy na podstawie opisu można jednoznacznie zidentyfikować rasę psa pokazanego na obrazku? Systemy identyfikacji obiektów mają praktyczne znaczenie, a ekspert musi czasami polegać tylko na opisie werbalnym, zadając pytania w celu uściślenia otrzymanego opisu. Ujednolicenie informacji zbieranej w raportach jest istotne, powstaje wiec pytanie, czy można ten proces skomputeryzować tak, by zadając kolejne pytania (podobnie jak w popularnej grze w 20 pytań) jednoznacznie określić rasę psa.

Najwięcej informacji można było znaleźć w języku angielskim, pozostawiam więc oryginalne nazwy. 329 ras psów przez nas użytych eksperci podzielili na 10 kategorii: Sheepdogs & Cattle Dogs; Pinscher & Schnauzer; Spitz & Primitive; Scenthounds; Pointing Dogs; Retrievers, Flushing Dogs & Water Dogs; Companion and Toy Dogs; Sighthounds. Szczegółowy opis dotyczy wyglądu: wielkości, wagi, koloru, rodzaju i długości sierści, kształtu uszu, pyska, ciała, ogona. Wiele informacji można znaleźć o trybie życia czy charakterze psów, ale są to cechy trudne do oceny przy krótkiej obserwacji (a całkiem niedostępne przy opisie zdjęcia). Pomimo wielu prób nie udało się automatycznie wygenerować informacji wystarczających do jednoznacznego rozpoznania rasy. Poprawianie i dopisywanie nowych informacji też wiele nie pomogło. Tradycyjna kategoryzacja opiera się na obserwacji zachowań, np. teriery wyglądają bardzo różnie, ale wszystkie kopią nory. Bez takiej informacji o danym psie nie da się go przypisać do kategorii teriery, w ramach której można poszukiwać konkretnej rasy.

Kategorie językowe i podobieństwo obrazów całkiem się rozmijają. Ontologie są słabo związane z podobieństwem wizualnym. Identyfikacja rasy jest za to możliwa na podstawie sylwetek uszu, głowy, pyska, łap, ciała i ogona, na podobnej zasadzie jak tworzenie portretów pamięciowych. Werbalny opis może aktywizować odpowiednią reprezentację ikonograficzną, przywołując z pamięci odpowiedni stan mózgu jeśli stan ten jest potencjalnie osiągalny, a więc został wcześniej zapamiętany. Kilka ras psów jest szeroko znanych więc użycie nazwy „jamnik" przekazuje w Europie właściwe wyobrażenie, ale już w większości krajów Azji tak nie będzie. Obraz można też starać się stworzyć od podstaw i zapamiętać, konstruując go z

istniejących elementów, odwołując się do podobieństwa do znanych obiektów. Durian ma kształt piłki rugby lub wielki zielony kasztan, ma ostre kolce. Modyfikacja znanych pojęć jest bardzo dobrym sposobem tworzenia nowych pojęć.

4. Symulacje komputerowe i czego się z nich możemy nauczyć

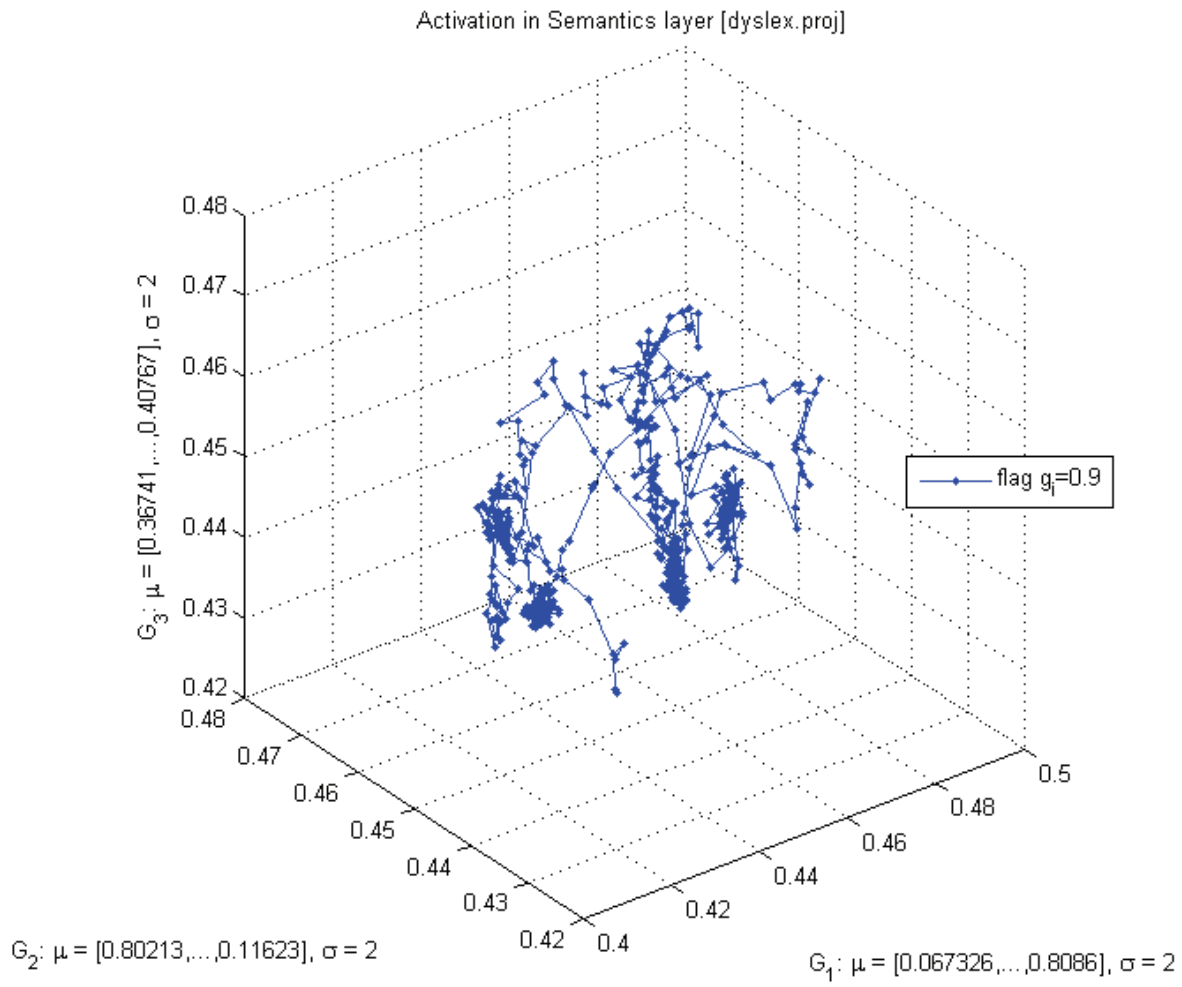
D1. Semantyka pojęć opiera się na aktywacji wielu obszarów mózgu. Nie wszystkim z nich potrafimy przypisać jednoznaczne funkcje. Aktywizacja części z nich, zwłaszcza obszarów znajdujących się w niedominującej (zwykle prawej) półkuli mózgu, nie da się opisać za pomocą pojęć mających sens, chociaż ich rola w ustalaniu tego sensu może być istotna. Rola ta może sprowadzać się do narzucania wielu ograniczeń na bieżąca interpretację tekstu lub obserwowanego zdarzenia, generowania antycypacji wynikających z przeżytych w przeszłości doświadczeń. Można więc mieć wątpliwości czy tak skomplikowane procesy da się obecnie modelować za pomocą sieci neuronowych. Jednakże nawet stosunkowo proste modele neurodynamiczne (czyli modele ilustrujące zmiany aktywności neuronów) pozwalają na pewne zrozumienie procesów neurobiologicznych związanych z językiem.

D2. Model opisany w poprzednim podrozdziale (punkt C2) został nauczony 40 słów, 20 słów abstrakcyjnych (rent, fact, deed, lack, gain …) i 20 konkretnych rzeczowników (tent, face, deer, hind, lock, rope, flag …). Uczenie polega na przypadkowym wyborze jednej z 3 warstw (ortografii, fonologii, lub semantyk) jako wejścia, a pozostałych dwóch jako wyjść na których chcemy otrzymać skorelowane wzorce, np. widząc słowo „flag” chcemy by w warstwie fonologicznej uaktywniły się w kolejnych kolumnach elementy fll@ggg, a w warstwie semantycznej uaktywniły się elementy kodujące mikrocechy charakteryzujące pojęcie „flag”, w stopniu zależnym od ich typowości. „Wyrób człowieka” (men-made) dobrze pasuje do słowa „flag”, ale „płeć” jest cechą, której nie można mu przypisać, za to pozwala odróżnić słowa „deer” (jeleń) od „hind” (łania). Wynikiem procesu uczenia jest taka zmiana parametrów sieci (głównie siły wzajemnych pobudzeń elementów sieci) by uzyskać pożądaną aktywność, odpowiednio skorelowaną we wszystkich warstwach. Wszystkie elementy (neurony) pomiędzy sąsiednimi warstwami jak i wewnątrz każdej warstwy są ze sobą połączone.

Model ten rozwinięto początkowo w celu symulacji głębokiej dysleksji, ale znalazł wiele zastosowań w symulacji procesów czytania. Wadą tego modelu jest wykorzystanie dość przy-

padkowych pojęć do jego uczenia. Ciekawszych rezultatów można się spodziewać używając większej liczby pojęć z jednej domeny.

D3. Ustalając aktywność elementów z jednej z warstw, np. ortograficznej, powodujemy rozchodzenie się aktywacji do sąsiednich warstw. Proces ten wymaga obliczania zmian aktywacji w kolejnych krokach czasu, prowadząc do narastania i zanikania aktywności różnych elementów, aż do przybliżonej stabilizacji po osiągnięciu zapamiętanego stanu. W języku służącym do opisu układów dynamicznych mówimy, że pojęcia są atraktorami dla trajektorii aktywności sieci neuronowej (lub ściślej, że tworzą obszary lub baseny atrakcji). Jeśli w chwili czasu t aktywność elementu o numerze i oznaczymy przez $a_i(t)$, to cały zbiór wartości tych aktywności można przedstawić jako punkt w przestrzeni aktywności a_i. Np. dla warstwy semantycznej mamy 140 elementów, więc przestrzeń ta będzie miała 140 wymiarów. Trajektoria aktywności opisana wektorem $\mathbf{A}(t)=\{a_i(t)\}$ tworzy w tej przestrzeni zbiór punktów obrazujący stan aktywności warstwy semantycznej. W pobliżu pewnych rozkładów aktywności elementów sieci zmiany stają się niewielkie i trajektoria skupia się wokół punktu, który nazywamy atraktorem punktowym. Punkt ten jest reprezentuje jakieś zapamiętane pojęcie, a patrząc na odpowiadający mu rozkład aktywności elementów warstwy semantycznej można określić jego własności.



Rys. 5 Trajektorie aktywności dla warstwy semantycznej po pobudzeniu słowem "flag".

D4. Wizualizacja trajektorii pokazuje, jak zmienia się z upływem czasu aktywność wszystkich użytych w symulacji neuronów (Dobosz i Duch, 2010, 2011). Każdy punkt na rys. 5 odpowiada określonej wartości aktywności 140 neuronów w danej chwili czasu. Ponieważ możemy widzieć jedynie 3 wymiary musimy dokonać takiej transformacji by z grubsza zachować wzajemne relacje pomiędzy kolejnymi punktami. Użyliśmy tu rozmytej dynamiki symbolicznej (Dobosz i Duch, 2010), ale techniczna strona jest mało istotna. Zagęszczenia trajektorii reprezentujące atraktory neurodynamiki pokazują, że system fluktuuje wokół danego pojęcia przez jakiś czas, a następnie przechodzi do pojęć skojarzonych, nie powraca jednak dokładnie do tego samego miejsca. Historia ewolucji systemu zmienia sytuację, pozostawiając we wszystkich warstwach ślady po poprzedniej aktywności.

Taki obraz jest wielkim uproszczeniem, gdyż zmiany ogólnego pobudzenia (efekty związane z emocjami, uwagą, zmęczeniem) mogą w znacznym stopniu zmienić dostępne stany i przebieg trajektorii. Spontaniczny proces skojarzeniowy powoduje, że po krótkim czasie przechodzimy od jednej myśli do drugiej, od pojęcia do pojęcia. Reprezentacja semantyczna danego pojęcia jest atraktorem neurodynamiki, czyli chwilowym spowolnieniem zmian aktywacji prowadzącym do kolejnych reprezentacji w procesie swobodnego kojarzenia. Na rysunku 5 widać początkową trajektorię zaczynająca się w środku, przy zerowych pobudzeniach elementów warstwy semantycznej, która po stosunkowo krótkim czasie wpada w obszar pierwszego atraktora, odpowiadający pojęciu *flag*. Jednakże po pewnym czasie system spontanicznie przechodzi do innego stanu, po niezbyt dużej liczbie kroków wpadając w kolejny atraktor związany z jakimś pojęciem. Oznacza to desynchronizację spójnej aktywacji neuronów, pojawienie się chaotycznego stanu przejściowego, w którym sieć próbuje znaleźć od nowa spójny podzbiór neuronów chętnych do współpracy, po czym pojawienie się nowego kwazistabilnego stanu.

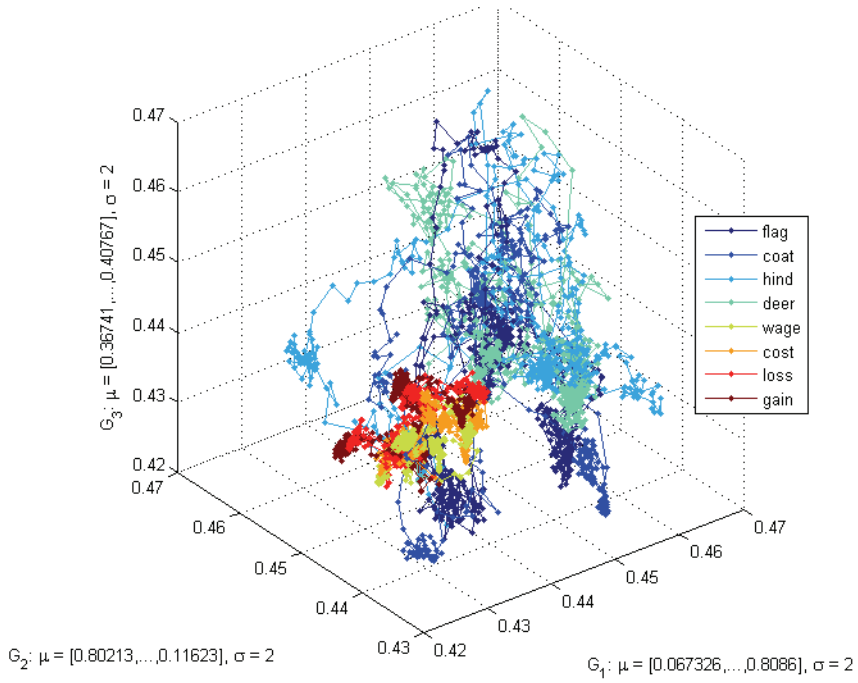
Dlaczego tak się dzieje, czy zawsze pojawia się reprezentacja znanego pojęcia i na jakiej zasadzie jest ono wybierane?

D5. W zależności od bieżącego stanu mózgu i historii ewolucji tego stanu trajektorie mogą się drastycznie zmienić, umożliwiając całkiem odmienne skojarzenia. Na stan neuronów wpływa wiele parametrów fizjologicznych: czasu ostatniego posiłku, przeżywane emocje, pobudzenie i zmęczenie, aktywność mózgu związana z bieżącymi czynnościami, percepcją i ruchem, skupianie uwagi, oraz efekty torowania wynikające z niedawnej aktywności (McNamara, 2005). Ma to wpływ zarówno na szybkość pojawiania się nowych skojarzeń jak to, jakie stany się uaktywnią. W danym momencie istnieje wiele potencjalnie osiągalnych stanów, które mogą stać się aktywne, ale mechanizmy konkurencji pomiędzy grupami neuronów, określane jako „zwycięzca bierze wszystko", hamują aktywność neuronów, które nie wchodzą w skład spójnej grupy charakteryzującej jedno pojęcie. To samo słowo w różnych zdaniach tworzy odmienne aktywacje, a jego leksykograficzne znaczenie zmienia się w sposób niemal ciągły w zależności od kontekstu. Informacje zawarte w tezaurusach muszą więc z konieczności być przybliżone, ograniczając się do najczęstszych, prototypowych znaczeń.

Prześledzenie wszystkich czynników, które decydują o powstaniu kolejnego skojarzenia nie jest możliwe, ale efekty dynamiczne związane z aktywacją pojęć są ostatnio intensywnie badane przez psycholingwistów (Spivey, 2007).

D6. Jeśli w warstwie semantycznej pojawi się aktywność elementów reprezentujących cechy charakteryzujące jakieś pojęcie pozostałe cechy, które je charakteryzują, również się pobudzą. W trzeciorzędowych obszarach kory dochodzi do integracji informacji z różnych modalności zmysłowych, pozwalając na identyfikację obiektów za pomocą wzroku, słuchu czy dotyku. Stany atraktorowe związane z pojęciami opisującymi te obiekty są osiągane na różnej drodze, tworząc wielowymiarowe reprezentacje odwołujące się do podobieństwa pomiędzy obiektami, ogólnymi cechami, które je charakteryzują. W szczególności cechy, które nie dają się skwantyfikować, takie jak specyficzne wrażenia (*qualia*) związane z pobudzeniem kory wzrokowej czy czuciowej, mają swój wkład powiększając obszary atrakcji i wpływając na wybór skojarzonych stanów, podczas gdy liczba związanych z nimi stanów w obszarach związanych z fonologia jest ograniczona. Język nie jest więc wstanie oddać bogactwa przeżyć wewnętrznych. Neurofenomenologia próbuje powiązać stany mózgu z wewnętrzna perspektywą subiektywnych przeżyć. Używając magnetycznej stymulacji przezczaszkowej można wyłączyć zlokalizowane obszary kory i dostrzec ich wkład w formowanie się takich stanów (Kim i inn, 2011), co stwarza intersujące możliwości badawcze.

Na rys. 6 przedstawiono stan warstwy semantycznej po pobudzeniu słowami *flag, coat, hind, deer*, odnoszącymi się do obiektów rzeczywistych (ciemniejsze linie), oraz *wage, cost, loss, gain*, odnoszącymi się do pojęć abstrakcyjnych (jaśniejsze linie). Reprezentacje tych pojęć są wyraźnie uboższe, charakteryzuje je mniej cech, gdyż nie pobudzają kory zmysłowej ani kory ruchowej, a więc związane z nimi aktywacje nie będą miały tak dużej wariancji jak aktywacje dla pojęć opisujących kategorie naturalne. Pojęcia matematyczne będą ściśle określone i powinny wykazywać minimalną wariancję.



Rys. 6 Trajektorie warstwy semantycznej dla 4 pojęć naturalnych i 4 abstrakcyjnych.

D7. Podobieństwa aktywacji mózgu są podstawą do tworzenia kategorii naturalnych. Skupienie silnie ze sobą skojarzonych aktywacji, pomiędzy którymi jest duże prawdopodobieństwo przejść (skojarzeń), tworzy rozmytą kategorię naturalną, do której można przypisać tą samą etykietę fonologiczną, czyli związać z nią słowo. Kategorie naturalne nie muszą mieć jednego zestawu cech je definiujących czy też jednego prototypu. Skojarzone stany związane z naturalnymi kategoriami mogą dotyczyć wyglądu, zachowania (np. terierów, psów kopiących nory) lub funkcji, możliwości działania, ruchu (np. „narzędzie"). Nie ma zbioru definiujących cech dla pojęcia „krzesło", ale różnorakie skojarzenia pozwalają nam rozpoznać dany obiekt jako krzesło. Pobudzenia neuronów odpowiedzialnych za semantykę kategorii naturalnych nie tworzą zbiorów wypukłych lecz mogą się składać z wielu rozłącznych reprezentacji (atraktorów), które należą do tej samej kategorii (mają wspólne reprezentacje fonologiczne).

D8. W modelu sieciowym (Rys. 2) możemy obserwować szybkie przejścia pomiędzy atraktorami. Dokładniejsza analiza (Dobosz i Duch, 2011) pokazuje, że mechanizm spontanicznego pojawiania się nowych myśli, widoczny w modelu jako przejście do kolejnego atraktora, związany jest ze „zmęczeniem" neuronów. Zbyt długi czas wysokiej aktywności uruchamia biologiczny mechanizm akomodacji, próg aktywacji potrzebny do ich pobudzenia wzrasta,

neurony mają tendencję do spontanicznej depolaryzacji i w efekcie przestają synchronizować swoją aktywność. Umożliwia to synchronizację innym neuronom, do tej pory hamowanych przez aktywność związaną z aktualnym stanem. Zaczynem ich synchronizacji stają się te neurony, które były słabo aktywne i jeszcze nie uległy zmęczeniu. Aktywność takich neuronów gwałtownie wzrasta i dołączają do nich inne neurony, dotychczas prawie nieaktywne. W danym momencie w mózgu wysoką aktywność (rzędu 40 impulsów na sekundę lub więcej) może wykazywać tylko około 1% neuronów, ale po chwili będzie to już inny 1% i w krótkim czasie prawie wszystkie neurony są silnie pobudzane.

Mechanizm akomodacji neuronów powoduje, że po okresie wysiłku umysłowego, skupienia nad tymi samymi pojęciami, spontanicznie pojawiają się całkiem odmienne myśli lub marzenia na jawie, luźno skojarzone z wcześniejszymi. Dzieje się tak zwłaszcza w wyniku ogólnego zmęczenia i braku energii. „Siła woli" to coś więcej niż tylko metafora. Glukoza potrzebna jest do wytwarzania neurotransmiterów i odżywiania neuronów, jej niski poziom powoduje utratę woli działania (Gailliot i Baumeister, 2007). Mózg człowieka w czasie pracy w jego obszarze ekspertyzy męczy się wolniej, bo więcej neuronów zaangażowanych jest w reprezentację pojęć, o których myśli.

Nowe stany nie zawsze odpowiadają nauczonym pojęciom, mogą być wynikiem synchronizacji fragmentów reprezentacji, tworząc nieistniejące pojęcia. Przypomina to mieszanie pojęć, stanowiące podstawę ogólnej teorii poznania Turnera i Fauconnier (2002). Pomimo niewielkiej liczby wyuczonych pojęć w naszej sieci zjawisko takie jest dość częste. Reprezentacje fonologiczne i ortograficzne pomagają ujednoznacznić pojawiające się stany, redukują wariancję pozwalając na ściślejsze określenie kategorii semantycznych. Po przeskoku do kilku skojarzonych pojęć stany atraktorowe coraz bardziej oddalają się od stanów wyuczonych. System może powrócić w pobliże początkowego stanu, ale nie będzie on dokładnie taki jak na początku. Dokładniejsze wnioski będzie można wyciągnąć po zbadaniu bardziej złożonego modelu nauczonego na większej liczbie pojęć z pojedynczej dziedziny.

D9. Procesy skojarzeniowe są stosunkowo proste i nie angażują w większym stopniu funkcji wykonawczych, sekwencyjnego wnioskowania w oparciu o chwilowo utrzymywane w pamięci roboczej informacje. Uczenie skojarzeń poprzez obserwację korelacji możliwe jest za pomocą prostej reguły Hebba: neurony jednocześnie aktywne powinny wzmocnić możliwości wzajemnego wpływu na siebie, tworząc podsieci sprzyjające synchronizacji. Uczenie się wy-

konywania zadań jest znacznie bardziej skomplikowane i wymaga uruchomienia mechanizmów nagrody, angażując wiele struktur podkorowych i płaty przedczołowe, w których powstają alternatywne plany działania.

Łatwo jest stworzyć model sieciowy potrzebny do generowania odpowiedzi na pytania z testu, w którym trzeba wybrać jedną z trzech odpowiedzi (O'Reilly i Munakata, 2000). Przypadkowy wybór powinien dać 1/3 poprawnych odpowiedzi, ale całkiem prosta sieć z jedną warstwą ukryta, która obserwuje korelacje pomiędzy słowami, osiąga w tym przypadku 60-80% poprawnych odpowiedzi, zależnie od wybranych pytań. Oczywiście taka sieć nie rozumie pytań ani odpowiedzi, nie jest zdolna nawet do najprostszego wnioskowania, a jej reprezentacja pojęć ogranicza się do korelacji pomiędzy występującymi z nią w jednym zdaniu słowami.

Jakiego rodzaju umiejętności należy testować w czasie egzaminu? Na pewno nie takie, które wymagają jedynie prostych skojarzeń.

D10. Modele sieciowe służą do przewidywań wpływu różnych uszkodzeń przepływu informacji w mózgu na zaburzenia funkcji językowych. Różne rodzaje dysleksji są rezultatem uszkodzenia połączeń pomiędzy warstwą ortograficzną, fonologiczną i semantyczną (O'Reilly i Munakata, 2000). W procesach demencji (np. w chorobie Alzheimera) utrata wiedzy o konkretnych pojęciach leżących stosunkowo nisko w ontologii następuje wcześniej niż dla pojęć ogólnych. W języku procesów neurodynamicznych możemy powiedzieć, że obszar atrakcji dla tych pojęć się kurczy, trajektorie nie są do nich przyciągane, gdyż zanikanie słabych ale licznych połączeń pomiędzy neuronami pozwala uaktywnić się tylko tym obszarom, które brały udział w kodowaniu wielu kategorii pojęć. Widać wyraźna korelację pomiędzy rozległością uszkodzeń (lezji) mózgu i poziomem w ontologii, na którym są utracone pojęcia. Im silniej dana cecha jest skorelowana z innymi tym dłużej jest użyteczna pomimo narastających uszkodzeń. Najpierw mylą się ze sobą nazwy zwierząt, a dopiero później nazwy obiektów mieszają się pomiędzy różnymi kategoriami (Taylor i inn, 2011; Devereux i inn, 2010).

D12. Parametry biofizyczne neuronu (zależne od jego molekularnej budowy, która wynika z procesów genetycznych i epigenetycznych sterujących budową komórek) decydują o tym jak szybko neurony się męczą. Ta obserwacja prowadzi do następującej hipotezy (Duch i inn. 2011): zbyt szybkie zmęczenie neuronów może doprowadzić do przeskakiwania od jednego

stanu umysłu do drugiego, w efekcie nie pozwalając na dłuższe skupienie się i stwarzając problemy z hiperaktywnością i deficytem uwagi (to jest przypadek ADHD), a zbyt słaba akomodacja neuronów prowadzi do skupienia się na jednym bodźcu i trudności w oderwaniu się od niego (tak jak w autyzmie). Obydwie choroby mogą więc być na dwóch krańcach tego samego spektrum. Badania genetyczne pokazują jedynie korelacje pomiędzy zmianami genetycznymi a zachowaniem (są one bardzo słabe w przypadku autyzmu i ADHD), ale nie określają mechanizmów, które są odpowiedzialne za zmiany zachowania. Tysiące różnych problemów na poziomie molekularnym może przejawiać się w podobny sposób w aktywności sieci, zaburzając mechanizmy synchronizacji neuronów. Badanie sieciowych modeli reprezentacji pojęć pokazują, jak zmiany parametrów neuronów wpływają na spontaniczne procesy kojarzeniowe.

D12. Rozumienie zdań wymaga szybkiego płytkiego wnioskowania potrzebnego do segmentacji mowy, analizy morfo-syntaktycznej, analizy składni, ujednoznacznienia sensu słów i ogólnej analizy semantyki, zrozumienia skojarzeń i analizy dyskursu. Cały ten proces przebiega bez wysiłku bardzo sprawnie, aktywność neuronalna w mózgu prowadzi do automatycznej interpretacji sensu zdania. Złożoność tego procesu widać najlepiej gdy próbuje się go zrealizować za pomocą symulacji komputerowych.

Z drugiej strony pomimo zaangażowania większości obszarów mózgu w procesy interpretacji pojęć i sensu zdania nawet stosunkowo proste konkluzje wymagające nietypowych skojarzeń mogą stwarzać wielkie trudności. Rozważmy takie 3 zdania:

- Wszyscy członkowie Akademii Magii to magicy.
- Żaden czarodziej nie jest członkiem Akademii Magii.
- Co konkretnie możesz powiedzieć o relacji między czarodziejami i magikami.

Nawet członkowie Mensy mają problem z wyciągnięciem prawidłowych konkluzji z tych dwóch zdań, a studenci na egzaminie podają kilkanaście błędnych odpowiedzi (Duch, 2010), twierdząc np. że bycie magikiem świadczy o tym, że nie jest się czarodziejem, lub że wszyscy magicy to czarodzieje. Wiele przykładów trudności w tworzeniu modeli mentalnych relacji logicznych pokazał Johnson-Laird rozwijając abstrakcyjną psychologiczną teorię modeli mentalnych (2006), ignorując procesy uczenia się i neurodynamikę mózgu. Jednakże bez komputerowych modeli tego procesu nasze rozumienie nie będzie pełne.

5. Dylemat plastyczności-stabilności.

E1. Sens pojęć zmienia się w czasie. Problem zmiany i stałości dyskutowany był od zarania filozofii. Kratylos w starożytnej Grecji głosił, że nawet znaczenie słów jest zmienne, dyskusja jest więc niemożliwa, bo zmienia się zarówno słuchacz jak i mówca. Na szczęście zmiany te są na tyle powolne, że dyskusja jest jednak możliwa, chociaż efektów związanych ze zmianą sensu pojęć nie można ignorować.

Kompromis pomiędzy stabilnością i plastycznością konieczny jest na wielu poziomach. Nasze możliwości poznawcze podlegają licznym ograniczeniom związanym z determinizmem genetycznym, neuronalnym i rolą czynników stochastycznych. Różnorodność kulturowa, wykształcenie i szeroki dostęp do informacji wpływają na plastyczność genetyczną i neuronalną, osłabiając efekty wpajania informacji w dzieciństwie. Zamknięte społeczeństwa sprzyjały silnemu determinizmowi neuronalnemu, tworząc sieci skojarzeń nie dopuszczające alternatywnych interpretacji, konstruując obraz świata odporny na zmiany.

E2. Szybkość ewolucji określona jest przez tempo mutacji DNA minus tempo napraw, przywracające stabilność genomu. Mutacje powstają dzięki promieniowaniu kosmicznemu, ultrafioletowemu, naturalnemu promieniowaniu tła, czynnikom chemicznym obecnym w środowisku i są częściowo naprawiane. Teoretycznie DNA mogło by się naprawiać bardziej sprawnie – zwierzęta żyjące na terenie wysokiego promieniowania w okolicach Czernobyla wykształciły takie mechanizmy – ale wtedy ewolucja nie zdołała by przystosować organizmów do katastrofalnych zmian warunków życia na Ziemi. Ceną ewolucji jest niedoskonałość mechanizmów naprawczych, pociągająca za sobą liczne choroby, od rzadkich chorób genetycznych w dzieciństwie do częstych chorób ujawniających się na starość.

E3. Zdolność do uczenia się zwiększa szanse przeżycia, ale nowa wiedza może silnie zaburzyć stabilny obraz świata: narzuca to ograniczenia na szybkość zmian (plastyczność) neuronów jak i na budowę całego mózgu. Zbyt duża stabilność neuronów oznacza brak adaptacji, uczenia się i zapamiętywania, a za duża plastyczność prowadzi do katastroficznego zapominania wcześniej zgromadzonej wiedzy, nowe fakty zniszczyłyby stabilny obraz świata, konieczny jest więc kompromis. Na poziomie synaptycznym wiąże się to z powolnymi zmianami sprawności przekazywania pobudzeń pomiędzy neuronami. Dokładne utrwalanie wszystkiego w pamięci bez konieczności powtarzania nie jest więc pożądane. Dlatego potrzebne są

liczne podsystemy pamięci, pozwalające na stopniową akumulację wiedzy z uwzględnianiem wyjątków. Ceną tego kompromisu są liczne niedoskonałości pamięci (Schacter 2003). Zmieniamy obraz świata utrwalony w pamięci semantycznej bardzo powoli – wyjątki „potwierdzają regułę", zamiast ją zdyskwalifikować, a obserwacje sprzeczne z naszymi przekonaniami są ignorowane lub zapominane. Potrafimy za to zapamiętać ważne epizody, które zdarzyły się tylko raz. Antycypacja przyszłości jest bardzo przydatna, ale wymaga daleko idącej ekstrapolacji rzeczywistej wiedzy, a to rzadko się udaje.

E4. Duże podobieństwo cech fizycznych i cech charakteru sprzyja spójności grupy, ale zmniejsza szanse na jej przeżycie w zmiennym, wrogim środowisku. W efekcie im większa zmienność w obrębie danego gatunku tym większa zdolność do przystosowania. *Homo sapiens* jest gatunkiem, który opanował wszystkie środowiska geograficzne, a ceną za to jest wielkie zróżnicowanie wewnątrzgatunkowe. Dotyczy to nie tylko wysokości czy wagi, ale i cech charakteru, altruizmu i egoizmu. W efekcie są zarówno święci jak i psychopaci.

Konieczny jest również kompromis pomiędzy opieraniem się na własnym doświadczeniu, a opieraniem się na wskazówkach i informacjach z drugiej ręki. Słuchanie autorytetów (np. słuchanie dorosłych przez dziecko) zwiększa szanse przeżycia, ale ślepa wiara w autorytety kiedy czasy się szybko zmieniają zmniejsza te szanse. Konieczne jest kwestionowanie autorytetów, poszukiwanie własnej drogi, dzięki czemu możliwe są zmiany. Jedni ludzie trzymają się poglądów wpojonych w dzieciństwie, inni stają się sceptyczni. Mity zwiększają spójność grupy uzasadniając wiele przydatnych tabu, jednakże zmiana warunków może unieważnić sens mitów. Szczególną rolę pełniły mity religijne, które nadal w wielu krajach są bardzo istotne. Państwa teokratyczne były bardzo stabilne, ale istniały na obszarach, gdzie zmiany klimatycznie były powolne a klęski żywiołowe rzadkie, nie było więc powodu do zmian. W takich warunkach strategia altruizmu odwzajemnionego jest opłacalna.

Efektem ubocznym słuchania autorytetów może być utrwalenie się „skrzywionego spojrzenia" na rzeczywistość, przesądów, uogólniania wyjątków, mylenia przypadkowych korelacji ze związkami przyczynowymi. Krótkotrwała wysoka plastyczność mózgu wywołana silnymi emocjami pozwala na „zamrożenie" wadliwego obrazu rzeczywistości, błędne skojarzenia i reakcje. Prowadzi to do błędów konfirmacji, tendencji do oceny rzeczywistości przez pryzmat swoich wcześniejszych przekonań, ignorowania informacji z nimi sprzecznych i trudności w formowaniu się nowych, bardziej realistycznych przekonań. Neuronalny determinizm zamyka

nas w klatce utrwalonych poglądów. Kora przedczołowa (PFC) pełni kluczową rolę w pamiętaniu wskazówek, a prążkowie w uczeniu się na podstawie obserwacji, które w przeszłości powiązane były z nagrodą. Obydwie struktury silnie reagują na dopaminę, której poziom regulowany jest przez kilka genów, mających różne warianty (Frank i inn. 2009; Doll i inn., 2011). To właśnie od nich zależy tendencja do trzymania się ustalonych reguł lub opierania się na własnych obserwacjach. Natura musi eksperymentować by w niesprzyjających warunkach ktoś przeżył: może to będą właśnie ci, którzy nie zmieniają zbyt łatwo poglądów, a może odwrotnie.

E5. Ciekawość wymaga podważenia powszechnie przyjętych wyjaśnień, prowadząc do postępu. Wywołane tym zmiany mogą prowadzić do niestabilności, nowe wynalazki (np. żelazo) wzmacniają chęć podbojów, wojen i związanych z nimi nieszczęść. Plagi chorób zabiły jednak znacznie więcej ludzi niż wojny, a dzięki ciekawości udało się wiele z tych chorób wyplenić. Konsekwencją strachu przed zmianami są ponure wizje w pełni stabilnego społeczeństwa totalitarnego, nowy, wspaniały świat opisany przez Aldousa Huxleya (1932).

Różnice cech charakteru powodują też różnicowanie się przekonań politycznych. Stabilność podkreślana jest przez partie konserwatywne, a konieczność większych zmian przez partie postępowe. Konserwatyzm sprzyja stabilności, zapobiegając zbyt szybkim zmianom prowadzącym do chaosu, liberalizm umożliwia adaptację do zmiennych warunków, ale może też odrzucić dobre rozwiązania, które pełniły istotną rolę społeczną. Zbyt konserwatywne kraje nie robią postępów, a zbyt postępowe wpadają w chaos. W krajach demokratycznych udaje się utrzymać pewien kompromis przez okresowe zmiany rządzącej partii, a w krajach o rządach autorytarnych dochodzi po dłuższym czasie do rewolucji.

E6. Natura lawiruje miedzy stabilnością a plastycznością, balansuje na krawędzi chaosu, z którego rodzą się nowe zjawiska. Świat zawsze zmienia się zbyt szybko dla starych i zbyt wolno dla młodych. Badania sposobu używania metafor u demokratów i republikanów w USA zgadzają się z ogólnymi przewidywaniami wynikającymi z rozumienia stabilności i plastyczności – republikanie są bardziej konserwatywni, a ich mózg są mniej plastyczne (Thibodeau, Boroditsky 2011). Różnice przejawiają się na poziomie genetycznym, w budowie mózgów, widoczne są nawet w sposobie kontrolowania ruchów sakadycznych oczu. Korelacje związków pomiędzy językiem, kulturą i zdolnościami poznawczymi są dość słabe, zależne od wielu cech osobowości (Greve, Wentura 2010).

6. Zakończenie

F1. Przedstawione powyżej rozważania oparte są na przekonaniu, że jedynie przez zrozumienie i aproksymację fizycznych stanów mózgu możemy dokonać istotnego postępu w rozumieniu natury pojęć i naszych własnych stanów mentalnych. Większość informacji analizowanej i przetwarzanej przez mózgi jest głęboko ukryta przed świadomym dostępem. Nie ma powodu by werbalny opis zachowania, a szczególnie tak skomplikowanej funkcji jaką jest posługiwanie się językiem naturalnym, był poprawny. U podstaw zachowania stoją bowiem procesy neurodynamiczne przebiegające w sposób ciągły. Takie procesy tylko w grubym przybliżeniu można opisać za pomocą skończonej liczby dyskretnych symboli (Spivey, 2007). Zachowanie nie jest przewidywalne, gdyż mózg jest plastyczny, ulega zmianie, a warunki reguł związane z możliwymi kontekstami nie dają się określić w skończonej liczbie sytuacji. Liczne normy i zachowania społeczne rozwinęły się po to by ograniczyć tą niepewność, wpływając na stabilizację zachowań człowieka, tworząc wzorce postępowania i narzucając mu ostre ograniczenia w postaci normatywnych reguł postępowania. Neuronalny determinizm zapobiegał chaosowi ale też ograniczał kreatywność i możliwości adaptacji, zwiększając bezwładność myślową.

F2. Nie mogliśmy dotychczas bezpośrednio obserwować tego, co dzieje się w mózgach, nie mamy doświadczenia z tego typu systemami, zmiany w nich zachodzące są szybkie i zależą od wielu czynników. Aparat pojęciowy psychologii potocznej nie może być tu adekwatny. Aktywność sieci neuronowych zmienia się w sposób trudny do przewidzenia. Nie zadajemy sobie sprawy z wielu złudzeń poznawczych (Piattelli-Palmarini, 1996; Pohl 2005), irracjonalnego myślenia (Ariely, 2008), mamy trudności z tworzeniem i analizą nawet stosunkowo prostych modeli mentalnych a zrozumienie wielu pojęć, np. abstrakcyjnego pojęcia liczby oderwanego od liczonych obiektów, zajęło setki lat. Opis pojęć za pomocą innych pojęć na tym samym lub wyższym poziomie abstrakcji nie odda w pełni natury procesów decydujących o ich rozumieniu i sposobach użycia. Jedynie aproksymacja zachodzących w mózgu rzeczywistych procesów neurodynamicznych odpowiedzialnych za procesy poznawcze daje szansę na zrozumienie całości procesów odpowiedzialnych za używanie pojęć, myślenie, zachowanie i podejmowanie decyzji.

F3. Idee ucieleśnienia i enaktywizmu są dla zrozumienia powstawania głębokich, percepcyjno-ruchowych reprezentacji podstawowych pojęć bardzo istotne (Barsalou, 2008), podkreślają konieczność wielomodalnych reprezentacji, ale to tylko fragment większej całości. Konieczne jest dodanie procesów wyższego rzędu, tworzenia się abstrakcyjnych reprezentacji pojęć w oparciu o obserwację relacji pomiędzy podstawowymi pojęciami (Mahon i Caramazza, 2006, Caramazza i Mahon, 2008). Z punktu widzenia transformacji pomiędzy stanami mózgu i powiązania neurodynamiki z procesami mentalnymi nie jest to trudne do wyobrażenia. Nie mamy jeszcze dobrych modeli komputerowych, ilustrujących przebieg takich procesów, nie wiemy dokładnie jak przebiega rozchodzenie się aktywności neuronalnej w mózgu. Doświadczenia prowadzone w ostatnich latach przy użyciu metod neuroobrazowania zbliżają nas do tego celu. Wizualizacja i uproszczony opis neurodynamicznych procesów pozwalający na obserwację ewolucji stanów mentalnych jest dobrą drogą do utworzenia pojęciowego opisu reprezentacji mentalnych odpowiadającego rzeczywistości.

F4. Lingwistyka neurokognitywna próbuje wykorzystać wiedzę o mózgu do zrozumienia procesów poznawczych, reprezentacji pojęć, zaburzeń neuropsychologicznych związanych z używaniem języka. Informatyka neurokognitywna jest nową dziedziną (Duch 2009), mająca na celu tworzenie praktycznych algorytmów czerpiących inspiracje ze zrozumienia procesów zachodzących w mózgu. Jej zastosowania obejmują analizę języka naturalnego (Duch i inn. 2008), kategoryzację pojęć w psychologii (Duch 1996a,1997) i nowe architektury kognitywne w sztucznej inteligencji (Duch, 2010a). Rozważania przedstawione w tym artykule można też rozszerzyć na zagadnienia dotyczące kreatywności (Duch i Pilichowski, 2007; Duch 2007a), oraz zrozumienie roli prawej półkuli mózgu w stanach Eureka, czyli nagłego wglądu pozwalającego na dostrzeżenie rozwiązania problemu (Bowden i inn. 2005; Jung-Beeman i inn. 2004; Duch, 2007).

F5. Zbudowanie matematycznego modelu umysłu, w którym można by rozpatrywać zdarzenia mentalne w powiązaniu z neurodynamiką jest nadal wielkim wyzwaniem. Z jednej strony mamy trudności z opisem doświadczenia wewnętrznego (Hurlburt i Schwitzgebel, 2007; Schwitzgebel, 2011). Z drugiej strony chociaż idee dotyczące geometrycznego opisu stanów mentalnych można łączyć z neurodynamiką dokonując transformacji mózg-umysł (Duch, 2010), to droga do stworzenia dokładnego modelu takich relacji jest daleka. Nie znamy szczegółów procesów zachodzących w mózgu, są trudności związane z badaniami ekspery-

mentalnymi, brak jest dobrych metod matematycznych do analizy sygnałów i procesów rozchodzenia się aktywacji w rzeczywistych sieciach neuronowych, jak i procesów przetwarzania informacji w oparciu o takie sieci. Pomimo tego rozproszone przetwarzanie informacji w sieciach neuronowych i koneksjonistycznych pozwala na modelowanie licznych funkcji poznawczych dając wyniki jakościowo porównywalne do rzeczywistych zachowań mózgu. Przedstawiony tu model czytania pozwala wyciągnąć wiele wniosków dotyczących zaburzeń procesu czytania, tworzenia się skojarzeń, dynamicznej natury samych pojęć i relacji semantyki z fonologią i ortografią. Być może dla otrzymania większości intersujących funkcji mentalnych nie trzeba bardzo szczegółowo modelować procesów zachodzących w mózgu, ważniejsza jest ogólna architektura (Duch 2005; Duch, Oentaryo, Pasquier 2008).

F6. Reprezentacji pojęć nie da się oddzielić od innych zagadnień, nie tylko związanych z językiem ale ogólnie z procesami myślenia i wnioskowania. Podejście neurokognitywne wprowadza tu nowy styl rozumowania ze specyficznymi problemami i pytaniami, daje nowy język opisu nieredukowalny do języka używanego dotychczas w lingwistyce czy psychologii. Dzięki symulacjom komputerowym i metodom eksperymentalnym można mieć nadzieję na znaczne postępy w zrozumieniu złożonych czynności poznawczych.

Podziękowania: Krzysztof Dobosz opracował program do wizualizacji atraktorów i przygotował rysunki trajektorii.

Literatura

Aisa, B., Mingus, B., O'Reilly, R. The emergent neural modeling system. Neural Networks, 21, 1045-1212, 2008.

Anderson M, Neural reuse: A fundamental organizational principle of the brain. Behavioral & Brain Sciences 33, 245–313, 2010.

Ariely D. (2008), Predictably irrational, The Hidden Forces That Shape Our Decisions. Harper-Collins.

Barsalou L.W. (2008), Grounded Cognition. *Annual Reviews of Psychology*, **59**, s. 617-645

Bowden, E.M., Jung-Beeman, M., Fleck, J. & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Science 9*, 322-328.

Brooks R. (1986), Elephants don't play chess. *Robotics and Autonomous Systems* **6,** s. 3-15.

Brooks R, Stein LA. (1994), Building Brains for Bodies. *Autonomous Robotics* **1,** s.7-25

Caramazza, A., & Mahon, B.Z. (2006). The organization of conceptual knowledge in the brain: the future's past and some future directions. Cognitive Neuropsychology, 23, 13-38

Churchland, P.M, A Neurocomputational Perspective: The Nature of Mind and the Structure of Science (1989);

Churchland, P.M, The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain, MIT Press, 1995.

Damasio H, Grabowski T.J., Tranel D., Frnak R.J, Hichiwa R.D, Damasio A.R. (1996): *A neural basis for lexical retrieval.* Nature 380: 499-505

Dehaene, S., Cohen, L. Sigman, M. & Vinckier, F. (2005) The neural code for written words: a proposal. *Trends in Cognitive Science 9*, 335-341.

Devereux, B., Pilkington, N., Poibeau, T., & Korhonen, A. (2010). Towards Unrestricted, Large-Scale Acquisition of Feature-Based Conceptual Representations from Corpus Data Research on Language and Computation, 7(2), 137-170

Dijksterhuis A, Nordgren L.F, A Theory of Unconscious Thought. Perspectives on Psych. Science 1(2), 95-109, 2006.

Dobosz K, Duch W. Understanding Neurodynamical Systems via Fuzzy Symbolic Dynamics. Neural Networks Vol. 23 (2010) 487-496, 2010

Dobosz K, Duch W, Visualization for Understanding of Neurodynamical Systems. Cognitive Neurodynamics 5(2), 145-160, 2011.

Doll B.B, Hutchison K.E, Frank M.J. Dopaminergic Genes Predict Individual Differences in Susceptibility to Confirmation Bias. Journal of Neuroscience, DOI: 10.1523/JNEUROSCI.6486-10.2011

Duch W (1994) A solution to the fundamental problems of cognitive sciences. UMK-KMK-TR 1/94, w International Philosophical Preprint Exchange.

Duch W (1996) From cognitive models to neurofuzzy systems - the mind space approach. Systems Analysis-Modelling-Simulation 24 (1996) 53-65

Duch W (1996a) Categorization, Prototype Theory and Neural Dynamics, 4th Int. Conference on SoftComputing'96, Iizuka, Japonia, str. 482-485.

Duch W. (1997), Platonic model of mind as an approximation to neurodynamics. W: *Brain-like computing and intelligent information systems*, red. S-i. Amari, N. Kasabov (Springer, Singapore), rozdz. 20, s. 491-512

Duch W. (2002), Geometryczny model umysłu. *Kognitywistyka i Media w Edukacji*, **6**, s. 199-230

Duch W. (2002a), Fizyka umysłu. *Postępy Fizyki* **53D**, s. 92-103

Duch, W. (2005). Brain-inspired conscious computing architecture. *Journal of Mind and Behavior 26(1-2)*, 1-22.

Duch W. (2007), Intuition, Insight, Imagination and Creativity. *IEEE Computational Intelligence Magazine* **2**(3), pp. 40-52.

Duch, W. (2007a) Creativity and the Brain. In: A Handbook of Creativity for Teachers. Ed. Ai-Girl Tan, Singapore: World Scientific Publishing, pp. 507-530.

Duch W. (2009), Neurocognitive Informatics Manifesto. W: Series of Information and Management Sciences, California Polytechnic State University, str. 264-282.

Duch W. (2010) Reprezentacje umysłowe jako aproksymacje stanów mózgu. Studia z Kognitywistyki i Filozofii Umysłu 3: 5-28, 2009

Duch W. (2010a), Architektury kognitywne. W: Neurocybernetyka teoretyczna, red. R. Tadeusiewicz, Wyd. Uniwersytetu Warszawskiego.

Duch W, Dobosz K, Attractors in Neurodynamical Systems. Advances in Cognitive Neurodynamics II (eds. R. Wang, F. Gu), pp. 157-161, 2011

Duch W, Pilichowski M. (2007). Experiments with computational creativity. *Neural Information Processing - Letters and Reviews 11*, 123-133.

Duch W, Matykiewicz P, Pestian J. (2008), Neurolinguistic Approach to Natural Language Processing with Applications to Medical Text Analysis. *Neural Networks* 21(10), 1500-1510.

Duch W, Nowak W, Meller J, Osinski G, Dobosz K, Mikołajewski D, and Wójcik G.M, Consciousness and attention in autism spectrum disorders. Proc. of Cracow Grid Workshop 2010, pp. 202-211, 2011.

Fauconnier G, Mental Spaces. Cambridge Uni. Press 1994

Fauconnier G, Mappings in Thought and Language Cambridge Uni Press, 1997.

Frank, M.J., Doll, B.B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nature Neuroscience, 12, 1062-1068, 2009.

Garagnani M, Wennekers T, Pulvermüller F (2009), Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network. Cognitive Computation 1(2):160-17

Gailliot, M. T., Baumeister, R. F. The physiology of willpower: Linking blood glucose to self-control. Personality and Social Psychology Review, 11, 303-327, 2007.

Gärdenfors P, Conceptual Spaces: The Geometry of Thought. MIT Press, 2000

Gigerenzer G. (2009), Intuicja. Inteligencja nieświadomości. Prószyński i Ska, Warszawa.

Gopnik A, Meltzoff A.N, Kuhl P.K, Naukowiec w kołysce. Wyd. Media Rodzina, 2004

Greve W., Wentura D., True lies: Self-stabilization without self-deception, Consciousness and Cognition 19(3), 721-730, 2010.

Harnad, S. (1990) The Symbol Grounding Problem. Physica D 42: 335-346.

Haynes, J.-D. & Rees, G. (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7, 523-534

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C. & Passingham, D. (2007). Reading hidden intentions in the human brain. *Current Biology* 17, 323-328.

Hurlburt R.T. Schwitzgebel E. (2007), Describing Inner Experience? Proponent Meets Skeptic. Cambridge, MA: MIT Press

Johnson-Laird, P.N . (1983). Mental models: Towards a cognitive science of language, inference and consciousness. Cambridge: Cambridge University Press.

Jung-Beeman, M., Bowden, E.M., Haberman, J., Frymiare, J.L., Arambel-Liu, S., Greenblatt, R., Reber, P.J., &. Kounios, J. Neural activity when people solve verbal problems with insight. *PLoS Biology 2*, 500-510, 2004.

Kim, J.G., Biederman I, Juan, C.H. The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: A TMS study. Journal of Neuroscience, 31, 8320-8324, 2011.

Lenat D. CYC: A Large-Scale Investment in Knowledge Infrastructure. Communications of the ACM 38(11), 33-38, 1995.

Mahon, B.Z, Caramazza, A. (2008). A Critical Look at the Embodied Cognition Hypothesis and a New Proposal for Grounding Conceptual Content. *Journal of Physiology - Paris*. **102**, s. 59-70

McNamara T.P, Semantic Priming. Perspectives from Memory and Word Recognition, Psychology Press 2005

Meunier, D., Lambiotte, R., Bullmore, E.T. (2010). Modular and hierarchically modular organization of brain networks. Frontiers in Neuroscience 4, 200, 1-11.

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. (2008), Predicting human brain activity associated with the meanings of nouns. *Science*. **30**;320(5880), s. 1191-95.

Monti M.M, Osherson D.N, Logic, Language and the Brain. Brain Research 2011 (w druku)

Monti M.M, Parsons L.M, Osherson D.N, The boundaries of language and thought: neural basis of inference making. PNAS 106(30), 12554-12559, 2009.

Okada, K, Hickok, G. (2006), Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *Neuroreport*, **17**, s. 1293-1296

O'Reilly R.C, Munakata Y. (2000), Computational Explorations in Cognitive Neuroscience Understanding the Mind by Simulating the Brain. Cambridge, MA: MIT Press

Piattelli-Palmarini M., Inevitable Illusions: How Mistakes of Reason Rule Our Minds (1996)

Pohl R., Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory (2005)

Pulvermuller, F. (2003), The Neuroscience of Language. On Brain Circuits of Words and Serial Order. Cambridge, UK:

Schacter D.L, Siedem grzechów pamięci. PIW 2003

Schwitzgebel E. Perplexities of Consciousness. MIT Press, 2011

Speer N.K, Reynolds J.R, Swallow K.M, Zacks J.M. Reading Stories Activates Neural Representations of Visual and Motor Experiences. Psychological Science 20(8): 989-999, 2009.

Spivey, M, The Continuity of Mind. Oxford University Press 2007

Taddeo, M., Floridi, L. (2005) The symbol grounding problem: A critical review of fifteen years of research. Journal of Experimental and Theoretical Artificial Intelligence, 17(4), 419-445.

Taylor, K.I., Devereux, B.J. & Tyler, L.K. Conceptual structure: Towards an integrated neuro-cognitive account. Cognitive Neuroscience of Language. W druku, 2011

Turner M, Fauconnier G: The Way We Think. Conceptual Blending and the Mind's Hidden Complexities. New York: Basic Books 2002

Varela, F., Thompson, E., & Rosch, E. (1991). The embodied mind: Cognitive science and human experience. Cambridge MA: MIT Press.

Von Glasersfeld E, Radical Constructivism: A Way of Knowing and Learning. London: Falmer Press 1995.

Zacks J.M, Speer N.K, Swallow K.M, Maley C.J. The brain's cutting-room floor: segmentation of narrative cinema. Frontiers in human neuroscience 4, 2010, 10.3389/fnhum.2010.00168.